

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 1999, 00:50:47 ; Search time 13.18 Seconds
(without alignments)
609.225 Million cell updates/sec

Title: US-09-072-994-14
Perfect score: 1791
Sequence: 1 MKLSDYYIDKELIYNSAISD.....IRYDRSKRITSKEILQLMLD 339

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1770	98.8	339	1 W99578	C. albicans Cdk act
2	276.5	15.4	349	1 W44767	MAP kinase. Schizo
3	261	14.6	544	1 W85028	CDK2-green fluores
4	261	14.6	544	1 W85029	Green fluorescent
5	261	14.6	298	1 W95689	Homo sapiens Cdc2
6	258.5	14.4	294	1 W95690	Oryza sativa Cdc2
7	255.5	14.3	424	1 R70768	JNK2 Oncoprotein p
8	255.5	14.3	416	1 W06795	High Osmolality G1
9	255.5	14.3	297	1 W95686	Schizosaccharomyce
10	254.5	14.2	423	1 R89409	Stress activated p
11	252.5	14.1	365	1 W55967	Human stress-activ
12	251.5	14.0	365	1 W23783	Human mitogen-acti
13	251.5	14.0	365	1 W26578	Human MAP kinase h
14	247.5	13.8	382	1 R90551	pJG4-5-CDK-BP clon
15	246.5	13.8	423	1 R89408	Stress activated p
16	246.5	13.8	365	1 W94073	Human CSBP beta po
17	246	13.7	317	1 W01559	CDK1 protein. New
18	246	13.7	317	1 W95682	Candida albicans C
19	243.5	13.6	375	1 W15512	MAP kinase #1. Wou
20	240.5	13.4	297	1 W95688	Homo sapiens Cdc2
21	238.5	13.3	297	1 W95687	Rattus norvegicus
22	236.5	13.2	384	1 R70767	JNK1 Oncoprotein p
23	236.5	13.2	633	1 W85009	Jnk1-green floures
24	236.5	13.2	630	1 W85018	Jnk1-green floures
25	235.5	13.1	323	1 W95685	Emericella nidulan
26	235	13.1	360	1 R71677	Human CSBP2. New c
27	235	13.1	360	1 W49825	Cytokine suppresss
28	235	13.1	360	1 W49066	Human monocyte CSA
29	235	13.1	607	1 W85010	p38-green flouresc
30	235	13.1	606	1 W85019	p38-green flouresc
31	234.5	13.1	411	1 R89412	Stress activated p
32	234	13.1	426	1 R89410	Stress activated p
33	234	13.1	384	1 R89411	Stress activated p
34	234	13.1	393	1 W61252	Salicylic acid ind
35	233	13.0	364	1 W40061	Human MAP kinase p
36	231	12.9	364	1 W65453	Drug binding prote
37	230	12.8	360	1 R71676	Human CSBP1. New c
38	230	12.8	360	1 W49824	Cytokine suppresss
39	230	12.8	360	1 W49065	Human monocyte CSA
40	227	12.7	423	1 W15513	MAP kinase #2. Wou
41	223.5	12.5	298	1 W95683	Saccharomyces cere
42	221.5	12.4	324	1 W95684	Ajellomyces capsul
43	218	12.2	393	1 R60700	Human ERK-5. Extra

44 218 12.2 379 1 W15506 Mitogen activating
45 218 12.2 631 1 W85006 Erk1-green floures

ALIGNMENTS

RESULT 1
W99578
ID W99578 standard; Protein; 339 AA.
AC W99578;
DT 22-JUN-1999 (first entry)
DE C.albicans Cdk activating kinase (CAK) in vivo (CIV) 1 protein.
KW protein kinase; cyclin-dependent kinase activating kinase; Cdk; CAK;
KW screening assay; fungicide.
OS Candida albicans.
PN WO9907836-A1.
PD 18-FEB-1999.
PF 11-AUG-1998; F01788.
PR 12-AUG-1997; FR-010287.
PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.
PA (CURI-) INST CURIE.
PA (CNRS) CNRS CENT NAT RECH SCI.
PI Faye G, Mann C, Thuret JY, Valay JG;
DR WPI; 99-180489/15.
DR N-PSDB; x27495.
PT Candida albicans protein with cyclin-dependent kinase-activating
PT kinase activity - used to screen compounds for fungicidal and
PT antimycotic activity
PS Claim 3; Page 12-13; 19; French.
CC This sequence represents a protein kinase of the CIV1 family that has
CC cyclin-dependent kinase (Cdk)-activating kinase (CAK) activity that is
CC not cyclin dependent. The new member designated CAK in vivo (CIV1)
CC differs from the SCCIV1 of Saccharomyces cerevisiae, and lacks the motif:
CC GXGX(Y/F)GXV, x = any amino acid. The sequence can be used screening
CC assays for fungicides, e.g. against Candida albicans, or as industrial
CC fungicides.
SQ Sequence 339 AA;

Query Match 98.8%; Score 1770; DB 1; Length 339;
Best Local Similarity 98.5%; Pred. No. 6.8e-150;
Matches 334; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKLSDYYIDKELIYNSAISDIYTAIDKFNLPVCLKIVDEDFSLPPHSIHREVLILKTLK 60
DB 1 MKLSDYYIDKELIYNSAISDIYTAIDKFNLPVCLKIVDEDFSLPPHSIHREVLILKTLK 60
QY 61 PHPNIIIEYFNLDKIYDDVILVTKLYRYDLSQLIEITKYCKRTRTFIYINGNLVSNQYTL 120
DB 61 PHPNIIIEYFNLDKICDDIILVTKLYRYDLSQLIEITKYCKRTRTFIYINGNLVSNQYTL 120
QY 121 ANEIEEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYDLK 180
DB 121 ANEIEEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYDLK 180
QY 181 LPPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEDIWLSGLITLTGLYSENFQSVLVKD 240
DB 181 LPPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEDIWLSGLITLTGLYSENFQSVLVKD 240
QY 241 DKELTNDSHVSDLYLLNQIFENFGTPNLTDDELFCDEYNENLHFKKFNLOKYPKDW 300
DB 241 DKELTNDSHVSDLYLLNQIFENFGTPNLTDDELFCDEYNENLHFKKFNLOKYPKDW 300
QY 301 DIILPRCNDLMKEIFTKMIYDRSKRITSKEILQLMLD 339
DB 301 DIILPRCNDLMKEIFTKMIYDRSKRITSKEILQLMLD 339

RESULT 2
W44767
ID W44767 standard; Protein; 349 AA.
AC W44767;

CC obtained by recording the variation, caused by the molecule, on
CC mechanically intact living cells, in the spatially distributed light
CC emitted from a luminophore present in the cells. The variation in light
CC emission is processed to provide information that correlates spatial
CC distribution to the degree of the molecule. The method is used to
CC identify agents that (in)directly affect intracellular signalling,
CC especially to screen for potential therapeutic agents or toxins, and to
CC identify new drug targets.
SQ Sequence 544 AA;

Query Match 14.6%; Score 261; DB 1; Length 544;
Best Local Similarity 26.2%; Pred. No. 2.4e-15;
Matches 83; Conservative 59; Mismatches 119; Indels 56; Gaps 11;
QY 21 IYTAIDKFNLPVCLKIV--DEDFSLPPHSIHREIFILKTLKPHNPNIIEYFNDLKIYDDV 78
Db 264 VYKARNKLTGEVVALKKIRLDTETEGVPSAIREISLLKELN-HPNIVKLLDVHTENKL 322
QY 79 ILVTKLYRYDLSQLIEITKYCKRTRTFIYGINGNLVSNQYTLANEIEEKDIKLWLSMSS 138
Db 323 YLVFEFLHQDLKKFMDASA-----LTGIPLPL-----IKSYLFQLLQ 359
QY 139 GLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFICYDLKLPKDEPPMAKYI-DVST 197
Db 360 GLAFCHSHRVLHRDLKPQNLLINTEGAIK--LADFGLARAFGV-----PVRTYTHEVVT 411
QY 198 GIYKAPELILGITYEYEDISLGIILTGLYSENFQSVLVKDDKELTNDSHVSDLYLLN 257
Db 412 LWYRAPEILLGSKYYSTAVDIWSLGCIF-----AEMVTRRALFPGDSEIDQLF 460
QY 258 QIFENFGTPNLTDFEDELFCDEYNNENLHFKFNKQYPRKDWDIILPRCNDLDMKEIFT 317
Db 460 RIFRTLGTDPDEVVPGVTSMPDY-----KPSFPKWARQDFSKVVPPLDED-GRSLLS 510
QY 318 KMIRYDRSKRITSKEIL 334
Db 511 QMLHYDPNKRISAKAAL 527

RESULT 5
W95689
ID W95689 standard; Protein; 298 AA.
AC W95689;
DT 08-JUN-1999 (first entry)
DE Homo sapiens Cdc2 protein kinase.
KW cdc2; protein kinase; diagnosis; infection; immunosuppression;
KW AIDS; acquired immunodeficiency syndrome; diagnosis; therapy;
KW organ transplantation.
OS Homo sapiens.
PN WO9856799-A1.
PD 17-DEC-1998.
PF 12-JUN-1998; U12100.
PR 13-JUN-1997; US-874347.
PA (MAYO-) MAYO FOUNDATION.
PI Gustafson MP, Leof EB, Limper AH, Thomas CF;
DR WPI; 99-080877/07.
PT New isolated polynucleotide encoding a Pneumocystis carinii Cdc2
PT polypeptide - useful for diagnosing or monitoring P. carinii
PT infection in patients with chronic immunosuppression
PS Example; Pages 43-44; 67pp; English.
CC The sequence is that of a Cdc2 polypeptide.
CC The Cdc2 polypeptides are useful in treating P. carinii
CC pneumonia in patients with chronic immunosuppression. Conditions
CC associated with P. carinii include acquired immune deficiency syndrome
CC (AIDS), solid tumors, hematological malignancies, organ transplantation,
CC and inflammatory conditions. The polypeptides are additionally useful
CC in identifying agents that inhibit the phosphorylation activity of
CC P. carinii Cdc2 polypeptides. The polypeptide and its antibodies are
CC useful for diagnosing or monitoring P. carinii infection in patients.
CC The P.carinii Cdc2 polypeptides allow a therapeutic approach to treating
CC P. carinii pneumonia because they are not limited by significant side
CC effects. Polynucleotides encoding p.carinii Cdc2 enable in vitro

CC production of the protein, which can be used in studying the organism
CC (especially its life cycle) and developing new therapies.
SQ Sequence 298 AA;
Query Match 14.6%; Score 261; DB 1; Length 298;
Best Local Similarity 26.2%; Pred. No. 1.1e-15;
Matches 83; Conservative 59; Mismatches 119; Indels 56; Gaps 11;

QY 21 IYTAIDKFNLPVCLKIV--DEDFSLPPHSIHREIFILKTLKPHNPNIIEYFNDLKIYDDV 78
Db 18 VYKARNKLTGEVVALKKIRLDTETEGVPSAIREISLLKELN-HPNIVKLLDVHTENKL 76
QY 79 ILVTKLYRYDLSQLIEITKYCKRTRTFIYGINGNLVSNQYTLANEIEEKDIKLWLSMSS 138
Db 77 YLVFEFLHQDLKKFMDASA-----LTGIPLPL-----IKSYLFQLLQ 113
QY 139 GLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFICYDLKLPKDEPPMAKYI-DVST 197
Db 114 GLAFCHSHRVLHRDLKPQNLLINTEGAIK--LADFGLARAFGV-----PVRTYTHEVVT 165
QY 198 GIYKAPELILGITYEYEDISLGIILTGLYSENFQSVLVKDDKELTNDSHVSDLYLLN 257
Db 166 LWYRAPEILLGSKYYSTAVDIWSLGCIF-----AEMVTRRALFPGDSEIDQLF 214
QY 258 QIFENFGTPNLTDFEDELFCDEYNNENLHFKFNKQYPRKDWDIILPRCNDLDMKEIFT 317
Db 214 RIFRTLGTDPDEVVPGVTSMPDY-----KPSFPKWARQDFSKVVPPLDED-GRSLLS 264
QY 318 KMIRYDRSKRITSKEIL 334
Db 265 QMLHYDPNKRISAKAAL 281

RESULT 6
W95690
ID W95690 standard; Protein; 294 AA.
AC W95690;
DT 08-JUN-1999 (first entry)
DE Oryza sativa Cdc2 protein kinase.
KW cdc2; protein kinase; diagnosis; infection; immunosuppression;
KW AIDS; acquired immunodeficiency syndrome; diagnosis; therapy;
KW organ transplantation.
OS Oryza sativa.
PN WO9856799-A1.
PD 17-DEC-1998.
PF 12-JUN-1998; U12100.
PR 13-JUN-1997; US-874347.
PA (MAYO-) MAYO FOUNDATION.
PI Gustafson MP, Leof EB, Limper AH, Thomas CF;
DR WPI; 99-080877/07.
PT New isolated polynucleotide encoding a Pneumocystis carinii Cdc2
PT polypeptide - useful for diagnosing or monitoring P. carinii
PT infection in patients with chronic immunosuppression
PS Example; Pages 44-45; 67pp; English.
CC The sequence is that of a Cdc2 polypeptide.
CC The Cdc2 polypeptides are useful in treating P. carinii
CC pneumonia in patients with chronic immunosuppression. Conditions
CC associated with P. carinii include acquired immune deficiency syndrome
CC (AIDS), solid tumors, hematological malignancies, organ transplantation,
CC and inflammatory conditions. The polypeptides are additionally useful
CC in identifying agents that inhibit the phosphorylation activity of
CC P. carinii Cdc2 polypeptides. The polypeptide and its antibodies are
CC useful for diagnosing or monitoring P. carinii infection in patients.
CC The P.carinii Cdc2 polypeptides allow a therapeutic approach to treating
CC P. carinii pneumonia because they are not limited by significant side
CC effects. Polynucleotides encoding P.carinii Cdc2 enable in vitro
CC production of the protein, which can be used in studying the organism
CC (especially its life cycle) and developing new therapies.
SQ Sequence 294 AA;

Query Match 14.4%; Score 258.5; DB 1; Length 294;

Best Local Similarity 25.3%; Pred. No. 1.7e-15;			
Matches 85; Conservative 60; Mismatches 132; Indels 59; Gaps 12;			
QY	5	DYYIDKELIYNSAISDIYTAIDKFNPLPVCLK---IYDEDFSLPPHSIHREIFILKTLKP	61
Db	2	EQYEKEEKIGEGFYGVYRARDKVTNETIALKKIRLEQDEGVSPFAT-REISLLKEMH-	60
QY	62	HPNIIEFYNDLKIYDDVILVTKLYRYDLSOLIE-ITKYCKRTRTFYINGNLVSNQYTL	120
Db	60	HGNIVRLHDVTHSEKRIYLVFEYLDLDLKKFMDSCPEFAKNPTL-----	104
QY	121	ANEIEEKDIKLWLSMSSGLEFIHSQGIHHRDIKPSNIFFARDDITQPI-IGDFDICYDL	179
Db	104	-----IKSYLYQILRGVAYCHSHRVLHRLDKPQNLLI--DRRTNALKLADFLARAF	153
QY	180	KLPPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEIDISLGIILGLYSENFQSVLYK	239
Db	154	GIPVRTFTH-----EVTLWYRAPEILLGSRQYSTPVDMSVGCIF-----AEMVN	199
QY	240	DDKELTNDSHVSDLYLLNQIFENFGTPNLTDFEDELFCDEYNNENLHFKFNLOKYPRKD	299
Db	200	QKPLFPGDSEIDELF--KIFRVLGTPNEQSWPGVSSLPDY-----KSAFFKWQAQD	248
QY	300	WDIILPRCNDLDMKEIFTKMIRYDRSKRITSKEILQ	335
Db	249	LATIVPTL-DPAGLDLLSKMLRYEPNKRITARQALE	283
RESULT 7			
R70768			
ID	R70768	standard; Protein; 424 AA.	
AC	R70768;		
DT	13-OCT-1995	(first entry)	
DE	JNK2	Oncoprotein polypeptide kinase.	
KW	Oncoprotein;	polypeptide kinase; JNK; c-Jun; AP-1; leukaemia;	
KW	prostate cancer;	lung cancer; colon cancer.	
OS	Homo sapiens.		
PN	W09503323-A.		
PD	02-FEB-1995.		
PF	18-JUL-1994;	U08119.	
PR	19-JUL-1993;	US-094533.	
PR	25-MAR-1994;	US-220602.	
PA	(REGC)	UNIV CALIFORNIA.	
PA	(UYMA-)	UNIV MASSACHUSETTS MEDICAL SCHOOL.	
PI	Davis R, Derijard B, Hibi M, Karin M, Lin A;		
DR	WPI; 95-075186/10.		
DR	N-PSDB; Q85309.		
PT	Oncoprotein protein kinase, JNK - useful for regulating c-Jun		
PT	activity and gene expression in cells		
PS	Claim 55; Figure 29;	143pp; English.	
CC	JNK1 (See R70767) and JNK2 phosphorylate the c-Jun N-terminal		
CC	activation domain which affect gene expression from AP-1 sites. The		
CC	2 protein kinases allow the detection of the level of specific		
CC	kinase activity associated with activation of c-Jun and AP-1. Cell		
CC	proliferative disorders associated with JNK can be treated by		
CC	administering a therapeutically effective amount of a reagent which		
CC	modulates JNK activity. Disorders such as leukaemia, cancers of		
CC	prostate, lung (non-small cell) and colon.		
SQ	Sequence	424 AA;	
Query Match 14.3%; Score 255.5; DB 1; Length 424;			
Best Local Similarity 27.1%; Pred. No. 5.3e-15;			
Matches 92; Conservative 49; Mismatches 112; Indels 87; Gaps 16;			
QY	21	IYTAIDKFNPLPVCLKIVDEDFSLPPHS--IHREIFILKTLKPHPNIIIEYFN-----DL	72
Db	40	VCAAFDTVLGISVAVKVKLSRPFQNTQAKRAYRELVLLKCVN-HKNIISLLNVFTPQKTL	98
QY	73	KIYDDVILVTKLYRYDLSOLIEITKYCKRTRTFYINGNLVSNQYTLANEIEEKDIKLW	132
Db	99	EEFQDVLYVLMELMDANLCQVIHM-----ELDHHERMSYL	131

QY	133	LKSMSSGLEFIHSQGIHHRDIKPSNIFFARDDITQPIIGDFDI---CYDLKLPPKDEPP	188
Db	132	LYQMLCGIKHLHSAGIIHRDLKPSNI-VVKS DCTKLIL-DFGLARTACTNFMMTPY----	186
QY	189	MAKYIDVSTGIYKAPELILGITNYEYEIDISLGIILGLYSENFQSVLVKDDKELTND	248
Db	186	-----VVTRYRAPEVILGM-GYKENVDIWSVGCIM-----GELVKGCVIFQGT	229
QY	249	HVSDLYLLNQIFENFGTPNLTDFEDELFCDEYNNENLHFKFNLOKYPR-----KD	301
Db	230	HIDQ---WNKVIEQLGTPS-AEFMKKIQPTVRNYVE-----NRPKYPGIKFEELFP	279
QY	302	IILP-----RCNDDLMKEIFTKMIRYDRSKRITSKEILQ	335
Db	279	-IFPSESERDKIKTSQARDLLSKMLVIDPDKRISVDEALR	317
RESULT 8			
W06795			
ID	W06795	standard; Protein; 416 AA.	
AC	W06795;		
DT	27-JAN-1997	(first entry)	
DE	High Osmolality Glycerol 1	gene product.	
KW	HOG1; hog1; hog2; HOG2;	mutant; glycerol; production; reduce; baking;	
KW	beverage; yeast; fermentation.		
OS	Saccharomyces cerevisiae.		
PN	US5545556-A.		
PD	13-AUG-1996.		
PF	23-MAY-1991;	704943.	
PR	23-MAY-1991;	US-704943.	
PR	18-MAR-1993;	US-032382.	
PA	(UYRI-)	UNIV RICE WILLIAM MARSH.	
PI	Gustin MC;		
DR	WPI; 96-383673/38.		
DR	N-PSDB; T44328.		
PT	yeasts contg. mutated HOG gene or multiple copies of HOG gene - have		
PT	reduced or increased glycerol production		
PS	Example 5; Fig 5;	28pp; English.	
CC	Low glycerol fermentation products can be produced using mutant strains		
CC	of yeast, e.g. Saccharomyces sp. or Torulaspora sp., that have a reduced		
CC	glycerol prodn. relative to the unmutated parental strain. High		
CC	Osmolality Glycerol genes have been identified, specifically HOG1, HOG2		
CC	and HOG4. Yeast strains contg. mutant HOG genes have a reduction in		
CC	glycerol biosynthesis and are useful in prodn. of novel beverages,		
CC	baked goods or glycerol. HOG1 and HOG2 are defined as yeast genomic DNA		
CC	segments capable of complementing the reduced growth in high osmolality		
CC	media (defined as the Osms phenotype) and glycerol biosynthesis defect		
CC	of the hog1 and hog2 mutants, respectively. Southern blot hybridisation		
CC	of HOG1 to electrophoretically separated S. cerevisiae chromosomes showed		
CC	hybridisation of HOG1 to chromosome XII. The present sequence is that of		
CC	the HOG1 gene product of 416 amino acids with a mol. wt. of 47 kDa.		
SQ	Sequence	416 AA;	
Query Match 14.3%; Score 255.5; DB 1; Length 416;			
Best Local Similarity 23.1%; Pred. No. 5.1e-15;			
Matches 84; Conservative 63; Mismatches 103; Indels 113; Gaps 13;			
QY	2	KLSDYYIDKELIYNSAISDIYTAIDKFNPLPVCLKIVDEDFS--LPPHSIHREIFILKTL	59
Db	18	EITNRYNDLNPVGMGAFGLVCSATDTLTSQPVAIKKMKPFSTAVLAKRYRELKLLKHL	77
QY	60	KPHPNII---EYFNDLKIYDDVILVTKLYRYDLSOLIEITKYCKRTRTFYINGNLVSN	116
Db	78	R-HENLICLDIF--LSPLEDIYFTELOGTDLHRLQ-----	113
QY	117	QYTLANEIEEKDIKLWLSMSSGLEFIHSQGIHHRDIKPSNIFFARDDITQPIIGDFDIC	176
Db	113	----TRPLEKQFVQYFLYQILRLGLKYVHSAGVIHRDLKPSNILIN-----CDLKIC	161
QY	177	YDLKLPPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEIDISLGIILGLYSENFQSV	236
Db	162	-DFGLARIQDPQMTGY--VSTRYVRAPEIMLTWQYDVEVDIWSAGCIF-----AE	209

QY 237 LVKDDKELTNDSHVSDLYLLNQIFENFGTNPNTDFEDELFCDEYNNENLHFKFNQLKYP 296
Db 210 MIEGKPLFPGKDHVHQSIIITDL---LGSP-----P 237
QY 297 RKDWIILPRCNDLMK-----EFTKMIRYDRSKRITSK 331
Db 238 K---DVINTICSENTLKFTVSLPHRDPPIPFSERFKTVPEPDAVLLLEKMLVDFPKKRITAA 294
QY 332 EIL 334
Db 295 DAL 297
RESULT 9
W95686
ID W95686 standard; Protein; 297 AA.
AC W95686;
DT 08-JUN-1999 (first entry)
DE Schizosaccharomyces pombe Cdc2 protein kinase.
KW cdc2; protein kinase; diagnosis; infection; immunosuppression;
KW AIDS; acquired immunodeficiency syndrome; diagnosis; therapy;
KW organ transplantation.
OS Schizosaccharomyces pombe.
PN WO9856799-A1.
PD 17-DEC-1998.
PF 12-JUN-1998; U12100.
PR 13-JUN-1997; US-874347.
PA (MAYO-) MAYO FOUNDATION.
PI Gustafson MP, Leof EB, Limper AH, Thomas CF;
DR WPI; 99-080877/07.
PT New isolated polynucleotide encoding a Pneumocystis carinii Cdc2
polypeptide - useful for diagnosing or monitoring P. carinii
infection in patients with chronic immunosuppression
PS Example; Page 41; 67pp; English.
CC The sequence is that of a Cdc2 polypeptide.
CC The Cdc2 polypeptides are useful in treating P. carinii
pneumonia in patients with chronic immunosuppression. Conditions
associated with P. carinii include acquired immune deficiency syndrome
(AIDS), solid tumors, hematological malignancies, organ transplantation,
and inflammatory conditions. The polypeptides are additionally useful
in identifying agents that inhibit the phosphorylation activity of
P. carinii Cdc2 polypeptides. The polypeptide and its antibodies are
useful for diagnosing or monitoring P. carinii infection in patients.
CC The P. carinii Cdc2 polypeptides allow a therapeutic approach to treating
P. carinii pneumonia because they are not limited by significant side
effects. Polynucleotides encoding P. carinii Cdc2 enable in vitro
production of the protein, which can be used in studying the organism
(especially its life cycle) and developing new therapies.
SQ Sequence 297 AA;
Query Match 14.3%; Score 255.5; DB 1; Length 297;
Best Local Similarity 24.3%; Pred. No. 3.3e-15;
Matches 81; Conservative 70; Mismatches 131; Indels 51; Gaps 11;
QY 7 YIDKELIYNSAISDIYTAIDKFNLPVCLK---IVDEDFSLPPHSIHREIFILKTLKPHP 63
Db 4 YQKVEKIGEGTYGVVYKARHKLSGRIVAMKIRLEDESEGVSTAI-REISLLKEVDEN 62
QY 64 NIEYFNDLKIYDDVILVTKLYRYDLSQLIEITKYCKRTRTFIYGINLVSNQYTLANE 123
Db 63 N---RSCNVRLLDILHAESKLYLVFEFLDMLKKYMDRISE-----TGATS 105
QY 124 IEEDKIKLWLSMSSGLEFIHSQGIHRDIKPSNIFFARDITQPIIGDFDICYDLKLPP 183
Db 106 LDPRLVQFTYQLVNGVNFCHSRRIIHRDLKPNLLIDKEGNLK--LADEGLARSGV-- 162
QY 184 KDEPPMAKYI-DVSTGIYKAPELILGITNYEYEIDISLGIILTGLYSENFQSVLVKDDK 242
Db 162 ---PLRNYTHEIVTLWYRAPEVLGSRHYSTGVDIWSVGCIF-----AEMIRSP 208
QY 243 ELTNDSHVSDLYLLNQIFENFGTNPNTDFEDELFCDEYNNENLHFKFNQLKYPKDWDI 302

Db 209 LFPGDSEIDEIF---KIFQVLGTNEEVPWPGVTLTLDYKSTFPRWKRMDLHK----- 258
QY 303 ILPRCNDLMKEIFTKMIRYDRSKRITSKEILQ 335
Db 258 VVPNGEEDAI-ELLSAMLVYDPAHRISAKRALQ 289
RESULT 10
R89409
ID R89409 standard; Protein; 423 AA.
AC R89409;
DT 26-APR-1996 (first entry)
DE Stress activated protein kinase p54 alpha-II.
KW Stress activated protein kinase; SAPK; p54 alpha-II; antibody;
KW inflammation; hypoxia; heat stress.
OS Rattus sp.
PN CA2148898-A.
PD 10-NOV-1995.
PF 08-MAY-1995; 148898.
PR 09-MAY-1994; US-240014.
PA (GEO) GEN HOSPITAL CORP.
PA (ONTA-) ONTARIO CANCER INST.
PI Avruch J, Banerjee P, Kyriakis JM, Woodgett JR;
DR WPI; 96-106355/12.
DR N-PSDB; T10639.
PT DNA encoding recombinant P54 stress activated protein kinase - and
related antibodies, useful for treating inflammation, hypoxia and
heat stress, and for drug screening
PS Claim 4; Page 35-36; 68pp; English.
CC A new family of proline-directed stress-activated protein kinases
CC (SAPK), designated p54 alpha-I, alpha-II, beta-I, beta-II and
gamma (R89408-12, respectively), are products of a set of cDNA
clones (see T10638-42) isolated from the brain cDNA library of
a cycloheximide-treated rat. The SAPK p54 proteins are strongly
activated by heat shock, tumour necrosis factor, interleukin-1-beta,
sphingomyelinase, chemical protein synthesis inhibitors and
ischaemia. They have potential utility in the modulation of the
inflammatory response and the up-regulation of repair or
protective cellular proteins following injury or chemical insult.
SQ Sequence 423 AA;
Query Match 14.2%; Score 254.5; DB 1; Length 423;
Best Local Similarity 27.1%; Pred. No. 6.4e-15;
Matches 92; Conservative 49; Mismatches 112; Indels 87; Gaps 16;
QY 21 IYTAIDKFNLPVCLKIYDEDFSLPPHS--IHREIFILKTLKPHNIEYFN-----DL 72
Db 40 VCAAFDTVLGINVAVKLSRPFQNTAKRAYRELVLKCVN-HKNIISLLNVFTPQKTL 98
QY 73 KIYDDVILVTKLYRYDLSQLIEITKYCKRTRTFIYGINLVSNQYTLANEIEEKDKLW 134
Db 99 EEFQDVYLYMELMDANLCQVIHM-----ELDHMSYL 131
QY 133 LKSMSSGLEFIHSQGIHRDIKPSNIFFARDITQPIIGDFDI----CYDLKLPPKDEPP 188
Db 132 LYQMLCGIKHLHSAGIHRDLKPSNI-VVKSDCITLIL-DFGLARTACTNFMTPY---- 186
QY 189 MAKYIDVSTGIYKAPELILGITNYEYEIDISLGIILTGLYSENFQSVLVKDDKELINDS 248
Db 186 -----VVTRYRAPEVILGM-GYKENVDIWSVGCIM-----GELVKGCVIFQGTD 229
QY 249 HVSDDLILLNQIFENFGTNPNTDFEDELFCDEYNNENLHFKFNQLKYP-----KWD 301
Db 230 HIDQ---WNKVIEQLGTPS-AEFMKKLOPTVRNYVE-----NRPKYPGIKFEELFPDW- 279
QY 302 IILP-----RCNDDLKKEIFTKMIRYDRSKRITSKEILQ 335
Db 279 -IFPSESERDKIKTSQARDLLSKMLVIDPKRISVDEALR 317
RESULT 11

PT Nucleic acid encoding mitogen-activated protein kinase homologue -
PT for production of recombinant protein, for treatment of
PT inflammation
PS Claim 1; Column 21; 19pp; English.24.
CC This polypeptide sequence comprises SNAP, a human MAP kinase
CC homologue that is involved in protective cell signalling processes.
CC Its amino acid sequence was deduced from a claimed isolated
CC polynucleotide (see T90352) obtained from a stomach cDNA library.
CC Also claimed are: an expression vector containing the polynucleotide
CC and a host cell transformed with the vector. SNAP and its variants
CC can be used in drug screening, drug design, research, antibody
CC production, and in methods for the diagnosis and treatment of
CC activated or inflamed cells and/or tissues associated with the
CC expression of SNAP, especially for the diagnosis and treatment of
CC conditions that affect the stomach such as gastritis, ulcers, viral
CC and bacterial infections and neoplasms.
SQ Sequence 365 AA;

Query Match 14.0%; Score 251.5; DB 1; Length 365;
Best Local Similarity 26.4%; Pred. NO. 9.8e-15;
Matches 91; Conservative 56; Mismatches 125; Indels 73; Gaps 14;

QY 2 KLSDDYIDKELIYNSAISDIYTAIDKFNLPVCLKIVDEDF--SLPPHSIHREIFILKTL 59
Db 20 ELPKTVSPTHVSGGAYGVCSAIDKRSGEKVAIKLSRPFQSEIFAIRAKRAYRELLLLKHM 79
QY 60 KPHPNII---EYF---NDLKIYDDVILVTKLRYDLSQLIEITKYCKRTTRFIYINGNL 113
Db 80 Q-HENVIGLLDVFTPASSLGNFYDFYLVMPFMQTDLQKI----- 118
QY 114 VSNQYTLANEIEEKDIKLMKSMSSGLEFIHSQGIHHRDIKPSNIFFARDITQPIIGDF 173
Db 118 -----MGMEFSEEKIQYLVYQMLKGLKYLHSGVVRDLKPGNL-AVNEDCELKIL--- 168
QY 174 DICYDLKLPKDEPPMAKIDVSTGIYKAPELILGINTNYEYEIDISLGLII---LTGLY 229
Db 168 ----DLGLARHADAEMTGY--VVTRWYRAPEVILSWMHYNQTVDIWSVGCIMAEMLTG-- 220
QY 230 SENFQSVLVKDDKELTNDSHVSDLYLLNQIFENFGTPTNLTDFEDELFCDEYNNENLHFKK 289
Db 220 -----KTLFKGKDYLDQ-----LTQILKVTGVPG-TEFVQKL-----NDKAASYI 259
QY 290 FNLQYPRKDWDIILPRCNDLMLKEIFTKMIYDRSKRITSKEIL 334
Db 260 QSLPQTPRKDFTLQFPSPQ-PADLLEKMLELDVDRKLTAAQAL 303

RESULT 14
R90551
ID R90551 standard; protein; 382 AA.
AC R90551;
DT 08-AUG-1996 (first entry)
DE pJG4-5-CDK-BP clone #227 derived CDK4 binding protein.
KW Cell cycle; CDK4; regulation; G1 phase; proliferation; tumourigenesis;
KW cyclin dependent kinase; differentiation; CDK4 inhibitor; agonist;
KW antagonist.
OS Synthetic.
PN WO9533819-A2.
PD 14-DEC-1995.
PF 02-JUN-1995; U07113.
PR 02-JUN-1994; US-253155.
PA (MITO-) MITOTIX INC.
PI Draetta G, Gyuris J;
DR WPI: 96-040227/04.
DR N-PSDB; T12180.
PT Cyclin-dependent kinase-4 binding protein - used in the isolation of
PT (ant)agonists of cell cycle regulation.
PS Claim 1; Page 87-88; 115pp; English.
CC R90533-R90556 are cyclin dependent kinase 4 (CDK4) binding proteins.
CC encoded by clones of the plasmid pJG4-5-CDKBP. CDK4 binding proteins
CC (CDK4-BP) may be used in an assay for screening test compounds as
CC inhibitors of CDK/CDK4-BP interaction. The complexes formed by CDK4

CC and D-type cyclins are strongly implicated in the control of the early
CC G1 phase of the cell cycle and are strong candidates for controlling
CC and/or preventing tumourigenesis and the onset of cancer. Nucleic acids
CC encoding CDK4-BP or fragments of these may be used as probes/primers
CC to diagnose the presence or absence of genetic lesions in a gene
CC encoding 1 of the 24 CDK4-BP, and hence to diagnose the risk for a
CC subject of developing a cell-proliferation associated disorder (e.g.
CC cancer).
SQ Sequence 382 AA;

Query Match 13.8%; Score 247.5; DB 1; Length 382;
Best Local Similarity 26.5%; Pred. NO. 2.4e-14;
Matches 89; Conservative 54; Mismatches 114; Indels 79; Gaps 17;

QY 21 IYTAIDKFNLPVCLKIVDEDFSLPPHS--IHREIFILKTLKPHPNIEYFN-----DL 72
Db 40 VCAAFDTVLGINVAVKLSRPFQNTAKRAYRELVLKCVN-HKNIISLNVFTPQKTL 98
QY 73 KIYDDVILVTKLRYDLSQLIEITKYCKRTTRFIYINGNLVSNQYTLANEIEEKDIKWL 132
Db 99 EEFQDVYLVLMELMDANLCQVIHM-----ELDHMSYL 131
QY 133 LKSMSSGLEFIHSQGIHHRDIKPSNIFFARDITQPIIGDFDI---CYDLKLPKDEPP 188
Db 132 LYQMLCGIKHLHSAAGIHRDLKPSNI-VKSDCTLKIL-DFGLARTACTNFMMPY---- 186
QY 189 MAKYIDVSTGIYKAPELILGINTNYEYEIDISLGLIITGLYSENFQSVLVKDDKELTND 248
Db 186 -----VVTRYRAPEVILGM-GYKENVDIWSVGCIM-----GELVKGCVIFQGT 229
QY 249 HVSDLYLLNQIFENFGTPTNLTDFEDEL--FCDEYNNENLHFKFNLQK-YPRKDWDIILP 305
Db 230 HIDQ---WNKVIEQLGTPS-AEFMKKQLQPTVRNYYENRPPKPGIKLELEFP--DW--LFP 281
QY 306 -----RCNDLMLKEIFTKMIYDRSKRITSKEILQ 335
Db 282 SESERDKIKTSQARDLLSQMLVIDPDKRISVDEALR 317

RESULT 15
R89408
ID R89408 standard; protein; 423 AA.
AC R89408;
DT 26-APR-1996 (first entry)
DE Stress activated protein kinase p54 alpha-I.
KW Stress activated protein kinase; SAPK; p54 alpha-I; antibody;
KW inflammation; hypoxia; heat stress.
OS Rattus sp.
PN CA2148898-A.
PD 10-NOV-1995.
PF 08-MAY-1995; 148898.
PR 09-MAY-1994; US-240014.
PA (GEHO) GEN HOSPITAL CORP.
PA (ONTA-) ONTARIO CANCER INST.
PI Avruich J, Banerjee P, Kyriakis JM, Woodgett JR;
DR WPI: 96-106355/12.
DR N-PSDB; T10638.
PT DNA encoding recombinant P54 stress activated protein kinase - and
PT related antibodies, useful for treating inflammation, hypoxia and
PT heat stress, and for drug screening
PS Claim 2; Page 34-35; 68pp; English.
CC A new family of proline-directed stress-activated protein kinases
CC (SAPK), designated p54 alpha-I, alpha-II, beta-I, beta-II and
CC gamma (R89408-12, respectively), are products of a set of cDNA
CC clones (see T10638-42) isolated from the brain cDNA library of
CC a cycloheximide-treated rat. The SAPK p54 proteins are strongly
CC activated by heat shock, tumour necrosis factor, interleukin-1-beta,
CC sphingomyelinase, chemical protein synthesis inhibitors and
CC ischaemia. They have potential utility in the modulation of the
CC inflammatory response and the up-regulation of repair or
CC protective cellular proteins following injury or chemical insult.
SQ Sequence 423 AA;

Query Match		13.8%;	Score 246.5;	DB 1;	Length 423;
Best Local Similarity		27.1%;	Pred. No. 3.3e-14;		
Matches 92;		Conservative 51;	Mismatches 110;	Indels 87;	Gaps 17;
QY	21	IYTAIDKFNLPVCLKIVDEDESLPPHS--IHREIFILKTLKPHNPNIIEYFN-----DL	72		
Db	40	VCAAFDTVLGINVAVKKLSRPFQNTAKRAYRELVLKCVN-HKNIISLLNVFTPOKTL	98		
QY	73	KIYDDVILVTKLYRYDLSQLIEITKYCKRTRTFIYGINGNLVSNQYTLANEIEEKDKLW	132		
Db	99	EEFQDVYLVMLMDANLCQVIHM-----ELDHERMSYL	131		
QY	133	LKSMSSGLEFIHSQGIHRDIKPSNIEFFARDDITQPIIGDFDI----CYDLKLPPKDEPP	188		
Db	132	LYQMLCGIKHLSAGIHRDLKPSNI-VVKSDCILKIL-DFGLARTACTNFMTPY----	186		
QY	189	MAKYIDVSTGIYKAPELIILGITNYEYEIDISLGIILTGLYSENFQSVLVKDDKELTND	248		
Db	186	-----VVTRYRAPEVILGM-CYKENVDIWSVGCIMA-----EMVL---HKSCSPGR	228		
QY	249	HVSDLYLLNQIFENFGTPNLTDFEDELFCDEYNNENLHFKFNLOKYP-----KDWD	301		
Db	229	DYIDQW--NKVIEQLGTPS-AEFMKKLOPTVRNVE-----NRPKYPGKFEELFPDW-	279		
QY	302	IILP-----RCNDDLMKEIFTKMIRYDRSKRITSKEILQ	335		
Db	279	-IFPSESERDKIKTSQARDLLSKMLVIDPDKRISVDEALR	317		

Search completed: November 6, 1999, 07:28:06
Job time: 23839 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 1999, 08:13:57 ; Search time 16.13 seconds
(without alignments)
1293.447 Million cell updates/sec

Title: US-09-072-994-14
Perfect score: 1791
Sequence: 1 MKLSDYYIDKELIYNSAISD.....IRYDRSKRITSKEILQLMLD 339

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTREMBL_10:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_invertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271.5	15.2	311	5	015851 leishmania
2	271.5	15.2	311	5	096526 leishmania
3	268	15.0	288	5	096821 plasmodium
4	264.5	14.8	407	3	059854 zygoscacchar
5	264.5	14.8	294	10	Q40484 nicotiana t
6	263.5	14.7	311	5	Q94796 trypanosoma
7	263.5	14.7	294	10	Q65838 lycopersico
8	262	14.6	391	11	Q61177 mus musculu
9	261	14.6	298	5	Q26671 theileria a
10	260	14.5	307	5	Q27168 paramecium
11	259.5	14.5	382	13	P79996 gallus gall
12	259	14.5	391	4	Q14013 homo sapien
13	258.5	14.4	366	5	Q62618 drosophila
14	257.5	14.4	380	3	093982 zygoscacchar
15	257.5	14.4	294	10	Q40482 nicotiana t
16	257.5	14.4	302	10	Q38772 antirrhinum
17	257	14.3	298	5	Q27032 theileria p
18	257	14.3	508	5	Q25917 plasmodium
19	256.5	14.3	765	5	Q25758 plasmodium
20	256.5	14.3	826	5	Q94656 plasmodium
21	255	14.2	307	10	Q04402 petunia hyb
22	255	14.2	369	10	Q43379 avena sativ
23	255	14.2	302	10	Q40734 oryza sativ
24	254.5	14.2	294	10	Q22292 allium cepa
25	253.5	14.2	294	10	P93101 chenopodium
26	253.5	14.2	371	10	Q04694 petroselinu
27	253	14.1	288	5	096820 plasmodium
28	252.5	14.1	365	4	Q15264 homo sapien
29	252.5	14.1	294	10	O82135 pisum sativ

30	252	14.1	346	11	Q60545
31	251.5	14.0	366	5	Q46216 mesocricetu
32	251.5	14.0	294	10	O65839 lycopersico
33	251	14.0	293	10	Q40483 nicotiana t
34	250	14.0	350	11	O54833 mus musculu
35	249	13.9	369	10	O81599 triticum ae
36	248	13.8	300	5	O44000 toxoplasma
37	248	13.8	377	5	Q17446 caenorhabdi
38	248	13.8	346	11	O55077 cricetulus
39	248	13.8	427	13	O42099 cyprinus ca
40	247.5	13.8	420	3	O59853 zygoscacchar
41	247.5	13.8	365	4	O14739 homo sapien
42	247.5	13.8	294	10	Q41639 vigna aconl
43	247	13.8	346	11	O55105 mus musculu
44	246.5	13.8	365	4	O15124 homo sapien
45	246.5	13.8	294	10	Q40789 petroselinu

ALIGNMENTS

RESULT 1
O15851
ID O15851; PRELIMINARY; PRT; 311 AA.
AC O15851;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TremBLrel. 10, Last annotation update)
DE CDC2-RELATED KINASE 3.
GN CRK3.
OS Leishmania mexicana.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98221140.
RA GRANT K.M., HASSAN P., ANDERSON J.S., MOTTRAM J.C.;
RT "The crk3 gene of Leishmania mexicana encodes a stage-regulated
cdc2-related histone H1 kinase that associates with p12.";
RL J. Biol. Chem. 273:10153-10159(1998).
DR EMBL; AJ001275; CAA04648.1; -.
DR PFAM; PF00069; pkinase; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
SQ SEQUENCE 311 AA; 35584 MW; 35C4E98D CRC32;

Query Match 15.2%; Score 271.5; DB 5; Length 311;
Best Local Similarity 26.9%; Pred. No. 1.9e-11;
Matches 94; Conservative 53; Mismatches 117; Indels 85; Gaps 14;

QY	5	DYIDKELIYNSAISDIYTAIDKFNLPVCLKIV--DEDFSLPPHSIHREIFILKTLKPH 62
Db	21	DRYNRLDVLGEGYGVYRAVDKITQYVALKKVRLDRTEEGIQTALREVSIQEF-DH 79
QY	63	PNIEYF----NDLKIYDDVILVTKLRYDLSQLIEITKYCKRTRFIYGINGLVSNQY 118
Db	80	PNIVNLLDVICSDGKLY---LVFEYVEADLKKAEKQE-----GGYSG----- 120
QY	119	TLANEIEEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICVD 178
Db	120	-----MDLKRLLYQLLDGLYFCHRRHRIHRDLKPNILLTSGNVLK--LADFGIARA 169
QY	179	LKLPPKDEPPMAKYY-DVSTGIYKAPELILGITNVEYIEDIWSLGIILTGLYSENFQSVL 237
Db	170	FQV-----PMHTYTHEVTVLWYRAPEILLGEKHYPAMDVMWSVGCIFAELAR---RKVL 220
QY	238	VKDDKELTNDSHVSDLYLLNQIFENFGTP-----NLTFEDELFCDEYNNENLH 286
Db	221	FRGDSEIGQ-----LFEIFQVLGTPTDTEGSWPGVSRLPDYRDV----- 260
QY	287	FKKFNLOQYPRKDWDIILPRCNDLDMKEIFTKMIRYDRSKRITSKEILQ 335
Db	260	-----FPRKWTAKRLGQVLPPELHPDAI-DLLSKMLKYDPRERISAKEALQ 302

```
RESULT 2
O96526
ID O96526 PRELIMINARY; PRT; 311 AA.
AC O96526;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CDC2-RELATED KINASE.
GN CRK3.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN V1;
RA WANG Y., DIMITROV K., GARRITY L.K., SAZER S., BEVERLEY S.M.;
RT "Stage-specific activity of Leishmania major CRK3 kinase and functional rescue of a Schizosaccharomyces pombe cdc2 mutant.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF073381; AAD08994.1; -.
KW PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
SQ SEQUENCE 311 AA; 35644 MW; C7716C68 CRC32;

Query Match 15.2%; Score 271.5; DB 5; Length 311;
Best Local Similarity 26.9%; Pred. No. 1.9e-11;
Matches 94; Conservative 53; Mismatches 117; Indels 85; Gaps 14;

QY 5 DYIDKELIYNSAISDIYTAIDKFNLPVCLKIV--DEDFSLPPPHSIHREIFILKTLKPH 62
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 21 DRYNRDLVLGEGYGVVYRAVDKITQYVVALKKVRLDRTEEGIPQATALREVSILQEF-DH 79
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 63 PNIIEYF---NDLKIYDDVILVTKLRYDLSQLIEITKYCKRTRTFIYGINGNLVSNQY 118
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 80 PNIVNLLDVICSDGKLY---LVFEYVEADLKKAEKQE-----GGYSG----- 120
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 119 TLANEIEEKDKLWLSKSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYD 178
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 120 -----MDLKRILYQLLDGLYFCHRHRIHRDLKPNILLTSGNVLK--LADFGLARA 169
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 179 LKLPKDEPPMAKYI-DVSTGIYKAPELILGITNYEYIDIWSLGIILTGLYSENFQSVL 237
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 170 FQV-----PMHTYTHEVWTLWYRAPEILLGEKHYPAYDMWSVGCIFAEL---TRRKVL 220
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 238 VKDDKELTNDSHVSDLYLLNQIFENFGP-----NLTDFEDELFCDEYNNENLH 286
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 221 FRGDSEIGQ-----LFEIFQVLGTPDTDEGSWPGVSRLPDYRDV----- 260
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 287 FKKFNLOQYPRKDWIILPRCNDLDMKEIFTKMIRYDRSKRITSKEILQ 335
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 260 -----FPKWTAKRLGOVLPELHPDAI-DLLSKMLKYDPRERISAKEALQ 302
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

RESULT 3
O96821
ID O96821 PRELIMINARY; PRT; 288 AA.
AC O96821;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CDC2-RELATED KINASE 2.
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LINE H;
RX MEDLINE; 96065755.
RA VINKENOOG R., VELDHUISEN B., SPERANCA M.A., DEL PORTILLO H.A.,
RA JANSE C.J., WATERS A.P.;
RT "Comparison of introns in a cdc2-homologous gene in a number of Plasmodium species.";
RL Mol. Biochem. Parasitol. 71:233-241(1995).
```

```
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LINE H;
RA VINKENOOG R., SPERANCA M.A., RAMESAR J., THOMAS A.W.,
RA DEL PORTILLO H.A., JANSE C.J., WATERS A.P.;
RT "Characterisation of the Cdc2-related kinase 2 gene from Plasmodium knowlesi and P. berghei.";
RL Mol. Biochem. Parasitol. 95:229-240(1998).
DR EMBL; AJ224155; CAA11852.1; -.
KW Kinase.
SQ SEQUENCE 288 AA; 32946 MW; 2E54A477 CRC32;

Query Match 15.0%; Score 268; DB 5; Length 288;
Best Local Similarity 27.5%; Pred. No. 2.9e-11;
Matches 84; Conservative 55; Mismatches 103; Indels 64; Gaps 11;

QY 35 LKIVDEDFSLPPPHSIHREIFILKTLKPPHPIEYFNDLKIYDDVILVTKLRYDLSQLIE 94
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 34 IRLEKEDEGIPSTAI-REISILKELK-HSNIVKLYDVHTKKRLILVFEHLDQDLKKLLD 91
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 95 ITKYCKRTRTFIYGINGNLVSNQYTLANEIEEKDKLWLSKSSGLEFIHSQGIHRDIK 154
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 92 V-----CDGGL-----ESVTAKSFLQLLSGIAYCHEHRVLRDLK 127
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 155 PSNIFFARDDITQPIIGDFDICYDLKLPKDEPPMAKYI-DVSTGIYKAPELILGITNYE 213
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 128 PQNLLINREGELK--IADFGLARAFGI-----PVRKYTHEVVTWLYRAPDILMGSKKYS 179
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 214 YEIDISLGIILTGLYSENFQSVLVKDDKELTND---SHVSDLYLLNQIFENFGPNTL 269
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 180 TPDIWSVGCIFA-----EMVNGRPLFPGVSETDQLMRIFRILGTPNSA 223
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 270 DFEDELFCDEYNNENLHFKFNLOQYPRKDWIILPRCNDLDMKEIFTKMIRYDRSKRIT 329
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 224 NWPSVTLPKYDPDFI-----VYEPLPWETFL-KGLDDTGIDLLSKMLRLDPNORIT 274
| | : : : | : | : | | | | | | | | | | | | | | | | | | |

QY 330 SKEILQ 335
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 275 AKEALQ 280

RESULT 4
O59854
ID O59854 PRELIMINARY; PRT; 407 AA.
AC O59854;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE ZHOGLP.
GN ZHOGL.
OS Zygosaccharomyces rouxii (Candida mogii).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Zygosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC42981;
RA WATANABE Y.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB012146; BAA25200.1; -.
DR PFAM; PF00069; pkinase; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
SQ SEQUENCE 407 AA; 46428 MW; 56BDD207 CRC32;

Query Match 14.8%; Score 264.5; DB 3; Length 407;
Best Local Similarity 23.7%; Pred. No. 7.8e-11;
Matches 86; Conservative 61; Mismatches 103; Indels 113; Gaps 13;

QY 2 KLSDYIIDKELIYNSAISDIYTAIDKFNLPVCLKIVDEFS--LPPHSIHREIFILKTL 59
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
Db 18 EITNRYTDLNPVGMGAFGLVCSATDTLAGQPVAIKKIMKPFSTAVLAKRTYRELKDLKHL 77
| | : : : | : | : | | | | | | | | | | | | | | | | | | |
```


OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
OC	euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;	
OC	core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;	
OC	Solanum.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CV. WEST VIRGINIA 106; TISSUE=FRUIT;	
RA	JOUBES J., CHEVALIER C., PHAN T.H., JUST D., ROTHAN C.,	
RA	BERGOUNIOUX C., RAYMOND P.;	
RL	Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.	
DR	EMBL: Y17225; CAA76700.1; -.	
DR	PFAM; PF00069; pkinase; 1.	
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.	
SQ	SEQUENCE 294 AA; 33751 MW; 75F8C00C CRC32;	
Query Match 14.7%; Score 263.5; DB 10; Length 294;		
Best Local Similarity 26.3%; Pred. No. 6.1e-11;		
Matches 89; Conservative 56; Mismatches 130; Indels 63; Gaps 13;		
QY	5 DYIDKELIYNSAISDIYTAIDKFNNLPVCLK---IVDEDFSLPPHSIHREIFILKTLKP 61	
Db	1 DQYEKVEKIGEGTYGVVYKARDRVNTEITALKKIRLEQDEGVSTAI-REISLLKEMQ- 60	
QY	62 HPNIEYFNLDKIYDDVILVTKLRYDLSQLIEITKYCKRTTRFIYGINGNLVSNQYTILA 121	
Db	60 HANIV-----RLQDVHSEKRLYLVEFYLDLDLKKHMSCEP----- 98	
QY	122 NEIEEKD---IKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPI-IGDFDICY 177	
Db	98 ----SKDPRLVKMFYQILRGIAVCHSHRVLHRDLKPQNLLI--DRRTNALKLADFGLAR 151	
QY	178 DLKLPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEIDWSLGIILTGLYSENFQSVL 237	
Db	152 AFGIPVRTFH-----EVTLWYRAPEILLGSRHSTPVDVWSVGCIF-----AEM 197	
QY	238 VKDDKELTNDSHVSDLYLLNQIFENFGTPNLTDFEELFCDEYNNENLHFKFNLQKYPR 297	
Db	198 VNQPPLFGDSEIDELF--KIFRVLTGTPNEDTWPGVTSLPDY-----KSAFPKWPP 246	
QY	298 KDWDIILPRCNDLMLKEIFTKMIRYDRSKRITSKEILQ 335	
Db	247 KDLAIVPNV-DGAGLDLLGKMLSLDPSKRITARNAL 283	
RESULT 8		
Q61177	PRELIMINARY; PRT; 391 AA.	
ID	Q61177	
AC	Q61177;	
DT	01-NOV-1996 (TReMBLrel. 01, Created)	
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)	
DT	01-NOV-1998 (TReMBLrel. 08, Last annotation update)	
DE	CASEIN KINASE II, ALPHA 1, RELATED SEQUENCE 4	
DE	(CASEIN KINASE II ALPHA SUBUNIT).	
GN	CSNK2A1-RS4.	
OS	Mus musculus (Mouse).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=B10.A/AKR;	
RA	CHEN L., HARDWICK J.P., SITKOVSKY M.V.;	
RL	Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.	
DR	EMBL; U51866; AAA96795.1; -.	
DR	MGD; MGI:894650; CSNK2A1-RS4.	
DR	PFAM; PF00069; pkinase; 1.	
SQ	SEQUENCE 391 AA; 45179 MW; 888B30D7 CRC32;	
Query Match 14.6%; Score 262; DB 11; Length 391;		
Best Local Similarity 26.5%; Pred. No. 1.1e-10;		
Matches 90; Conservative 59; Mismatches 125; Indels 66; Gaps 13;		

QY	5 DYIDKELIYNSAISDIYTAIDKFNNLPVCLKIVDEDFSLPPHSIHREIFILKTLKPHN 64	
Db	38 DYQLVRKL-GRGYSEVFEAINITNNEKVVVKILK---PVKKKKIKREIKILENLRGGPN 93	
QY	65 IIEYFNLDKIYDDVILVTKLRYDLSQLIEITKYCKRTTRFIYGINGNLVSNQYTLANEI 124	
Db	94 IITLADIVK--DPV-----SRPALVFEHVNTDFKQ--LYQTL 128	
QY	125 BEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYDLKLPPK 184	
Db	129 TDYDIRFYMEILKALDYCHSMGIMHRDVKPHNVIMIDHEHRKLRLI-----DWGLAEF 181	
QY	185 DEPPMAKYIDVSTGIYKAPELILGITNYEYEIDWSLGIILTGLY--SENFQSVLVKDDK 242	
Db	182 YHPGQYNNVRVASRYFKGPELLVDYQMYDYSLDMWSLGCMLASMIFRKEPF----- 233	
QY	243 ELTNDSHVSDLY-LLNQIFENFGTPNLTDFEELFCDEYNNENLHFKFN--LQKYPKRD 299	
Db	233 -----FHGHDNYDQLVRIAKVLGTEDLYD-----YIDKYNIE--LDPRENDILGRHSKR 280	
QY	300 WDIILPRCNDLMLK---EIFTKMIRYDRSKRITSKEILQ 335	
Db	281 WERFVHSENQHLVSPALDFLDKLLRYDHQSRLTAREAME 320	
RESULT 9		
Q26671	PRELIMINARY; PRT; 298 AA.	
ID	Q26671	
AC	Q26671;	
DT	01-NOV-1996 (TReMBLrel. 01, Created)	
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)	
DT	01-NOV-1998 (TReMBLrel. 08, Last annotation update)	
DE	CDC2-LIKE KINASE.	
OS	Theileria annulata.	
OC	Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;	
OC	Theileria.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ANKARA;	
FX	MEDLINE; 97084571.	
RA	KINNAIRD J.H., LOGAN M., KIRVAR E., TAIT A., CARRINGTON M.;	
RT	"The isolation and characterization of genomic and cDNA clones coding	
RT	for a cdc2-related kinase (ThCRK2) from the bovine protozoan parasite	
RT	Theileria."	
RL	Mol. Microbiol. 22:293-302(1996).	
DR	EMBL; X98768; CAA67306.1; -.	
DR	PFAM; PF00069; pkinase; 1.	
SQ	SEQUENCE 298 AA; 34230 MW; BF4362AE CRC32;	
Query Match 14.6%; Score 261; DB 5; Length 298;		
Best Local Similarity 24.9%; Pred. No. 9.1e-11;		
Matches 85; Conservative 71; Mismatches 109; Indels 76; Gaps 17;		
QY	7 YIDKELIYNSAISDIYTAIDKFNNLPVC---LKIVDEDFSLPPHSIHREIFILKTLKPH 62	
Db	4 YHKMEKIGEGTYGVVYKA--QNNHGEICALKKIRVEEDEGIPSTAI-REISLLKELH-H 59	
QY	63 PNIEYFNLDKIYDDVILVTKLRY---DLSQLIEITKYCKRTTRFIYGINGNLVSNQYT 119	
Db	60 PNIV-WLRDV-IHSEKCL-TLVEFYLDQDLKLLD-----ACDGG----- 98	
QY	120 LANEIEEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYDL 179	
Db	98 -----EPTTAKSFLYQILRGISYCHDHRILHRDLKPQNLLINREGVLK--LADFGLARAF 150	
QY	180 KLPPKDEPPMAKYI-DVSTGIYKAPELILGITNYEYEIDWSLGIILTGLYSENFQSVLV 238	
Db	151 AI-----PVRSYTHEVWTLWYRADPVLMSKKYSTAVDIWSVGCIFA----- 193	
QY	239 KDDKELTND---SHVSDLYLLNQIFENFGTPNLTDFEELFCDEYNNENLHFKFNLQK 294	
Db	193 ----EMINGVPLFPGISEQDLKRIKILGTPNVDVSNPQVNVNLPAYNPDFCYEK----- 244	

QY 295 YPRKWDIILPRCNDLDMKEIFTKMIYDRSKRITSKEILQ 335
: | | | : : : : : | | | : : : : : | | | :
Db 244 ---QAWSSIVPKLINESGI-DLISRLMLQDLPVQRIASAKALK 280

RESULT 10
Q27168
ID Q27168 PRELIMINARY; PRT; 307 AA.
AC Q27168;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)
DE CDC2.
GN CDC2.

OS Paramaecium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Nassophorea; Peniculida;
OC Paramaecium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=51-S;
RX MEDLINE; 95210349.
RA TANG L., PELECH S.L., BERGER J.D.;
RT "Isolation of the cell cycle control gene cdc2 from Paramaecium
tetraurelia";
RL Biochim. Biophys. Acta 1265:161-167(1995).
DR EMBL; U15802; AAA79977.1; -.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 307 AA; 35305 MW; EIC77C7C CRC32;

Query Match 14.5%; Score 260; DB 5; Length 307;
Best Local Similarity 26.6%; Pred. No. 1.1e-10;
Matches 89; Conservative 64; Mismatches 122; Indels 60; Gaps 13;

QY 2 KLSDYIIDKELIYNSAISIYTAIDKFNLPVCLK--IVDEDFSLPPHSIHREIFILKT 58
|| | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 5 KLEKEYEKEKLGEGYI--VYKALDRNTNEYVAIKKIRLESEBEGIPSTAI-REISLLKE 61
QY 59 LKPHNIIIEYFNDLKIYDDVILVTKLYRYDLSQLIEITKYCKRTTRFIYINGNLVSNQY 118
| | | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 62 LN-HPNIVKLMEVHSNKKLVLFVEYFEMDLKKFLA-----QF 98

QY 119 TLANEIEEKDIKMLKSSGLEFIHSGIHRDIKPSNIFFARDDITQPIIGDFICYD 178
: | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 99 PKEGMEPVIVKSFYQLLRGIQACHQKILHRDLKPQLNLGSDGILK--LADFGLARA 156
QY 179 LKLPKDEPPMAKYIDVSTGIYKAPELIGITNIEYEIDISLGIILTGLYSENFQSVLV 238
: | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 157 SGIPVKSFTH-----EVTWLWYRPPDVLGSKNYNTSIDISVGC-----FGE-----M 201
QY 239 KDDKELTNDSHVSDLYLLNQIFENFGTPTNLTDFEDELFCDEYNNENL-HFKKFNLOKY-P 296
: | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 202 SNLKPLFAGSNETD--QLKKIFRVLTGTPSPIEYPKLNDLPSWKPENFEQYQPDNLAKFCP 259

QY 297 RKDWDIILPRCNDLDMKEIFTKMIYDRSKRITSK 331
| | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 260 RLDPDGL-----DLLVKMLKINPDQITAK 284

RESULT 11
P79996
ID P79996 PRELIMINARY; PRT; 382 AA.
AC P79996;
DT 01-MAY-1997 (TremBLrel. 03, Created)
DT 01-MAY-1997 (TremBLrel. 03, Last sequence update)
DT 01-JAN-1999 (TremBLrel. 09, Last annotation update)
DE C-JUN AMINO-TERMINAL KINASE-2 ALPHAL.
GN JNK2-ALPHA1.

OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria; Aves;
OC Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE; 97320448.
RA ISHIKAWA T., NAKADA-MORIYA Y., ANDO C., TANDA N., NISHIDA S.,
RA MINATOGAWA Y., NOHNO T.;
RT "Expression of the JNK2-alpha1 gene in the developing chick brain.";
RL Biochem. Biophys. Res. Commun. 234:489-492(1997).
DR EMBL; AB000807; BAA19188.1; -.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 382 AA; 44009 MW; 39048F6D CRC32;

Query Match 14.5%; Score 259.5; DB 13; Length 382;
Best Local Similarity 27.9%; Pred. No. 1.6e-10;
Matches 95; Conservative 49; Mismatches 109; Indels 87; Gaps 17;

QY 21 IYTAIDKFNLPVCLKIVDEDFSLPPHS--IHREIFILKTLKPHNIIIEYFN-----DL 72
: | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 40 VCAAFDTVLGINVAVKLSRPFQNTAKRAYRELVLKCVN-HKNIISLLNVFTPKSL 98
QY 73 KIYDDVILVTKLYRYDLSQLIEITKYCKRTTRFIYINGNLVSNQYTLANEIEEKDIKLW 132
: | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 99 EEFQDVYLVMELDANLCQVIHM-----ELDHMERMSYL 131
QY 133 LKSMSSGLEFIHSGIHRDIKPSNIFFARDDITQPIIGDFDI-----CYDLKLPPKDEPP 188
| | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 132 LYQMLCGIKHLHSAGIHRDLKPSNI-VVKSDDLTKIL-DFGLARTACTNFMTPY----- 186
QY 189 MAKYIDVSTGIYKAPELIGITNIEYEIDISLGIILTGLYSENFQSVLVKDDKELTND 248
| | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 186 -----VTRYRAPEVILGM-GYKENVDIWSVGCIM-----GELVKGCVIFQGT 229
QY 249 HVSDLYLLNQIFENFGTPTNLTDFEDELFCDEYNNENLHFKKFNLOKYPR-----KDWD 301
: | | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 230 HIDO---WNKVIEQLGTPS-AEFMKKQLQPTVRNYVE-----NRPKYPGIKFEELFPDW- 279
QY 302 IILPRCND-DLMK-----EFTKMIYDRSKRITSKEILQ 335
| | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 279 -IFPSESDDRDKLKTQARDLLSKMLVDPDKRISVDEALR 317

RESULT 12
Q14013
ID Q14013 PRELIMINARY; PRT; 391 AA.
AC Q14013;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TremBLrel. 08, Last annotation update)
DE CASEIN KINASE II ALPHA SUBUNIT.
GN CKII ALPHA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93131026.
RA DEVILAT I., CARVALLO P.;
RT "Structure and sequence of an intronless gene for human casein kinase
II-alpha subunit.";
RL FEBS Lett. 316:114-118(1993).
DR EMBL; X70251; CAA49758.1; -.
DR PFAM; PF00069; pkinase; 1.
SQ SEQUENCE 391 AA; 45156 MW; 3FFCA0AD CRC32;

Query Match 14.5%; Score 259; DB 4; Length 391;
Best Local Similarity 26.5%; Pred. No. 1.7e-10;
Matches 90; Conservative 58; Mismatches 126; Indels 66; Gaps 13;

QY 5 DYITDKELIYNSAISIYTAIDKFNLPVCLKIVDEDFSLPPHSIHREIFILKTLKPHN 64
| | : : : : : | | | : : : : : | | | : : : : : | | | : : : : :
Db 38 DYQLVRKL-GRGKYSEVFEAINITNNEKVVVKILK---PVKKKKKIKREIKILENLRGPN 93

OC core eudicots; Asteridae; euasterids I; Solanales; Solanaceae;
OC Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=XANTHI;
RA QIN L.X., PERENNES C., BERGOUNTOUX C.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; L77082; AAB02567.1; -.
DR MENDEL; 13986; Nicotiana glauca; mn13986.
DR PFAM; PF00069; pkinase; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
SQ SEQUENCE 294 AA; 33874 MW; 9B0479D4 CRC32;

Query Match		14.4%	Score 257.5;	DB 10;	Length 294;
Best Local Similarity		25.9%	Pred. No. 1.6e-10;		
Matches 90;		Conservative 54;	Mismatches 122;	Indels 81;	Gaps 14;
QY	5	DYIDKELIYNSAISDIYTAIDKFNLPVCLK--IVDEDFSLPPHSIHREIFILKTLKP	61		
Db	2	DQYEKVEKIGEGTYGVVYKARDRVNTETIAKKIRLEQDEGVPSTAI-REISLLKEMQ-	60		
QY	62	HPNIIIEYFNDLKIYDDVILVTKLYRYDLSQLIEITKYCKRTTRFIYGINGNLVSNQYTLA	121		
Db	60	HANIV-----RLQDVVHSEKRLYLVEYLDLDLKKHMDSSPEF-----	98		
QY	122	NEIEEKD---IKLWLKSMSSGLEFIHSQGIHHRDIKPSNIEFFARDDITQPI-IGDFDICY	177		
Db	98	-----SKDPRLVKMFYQILRGIAVCHSHRVLHRDLKPQNLI--DRRTNALKLADFGLAR	151		
QY	178	DUKLPPKDEPPMAKYIDVSTGIYKAPELILGITNVEYEIDWSLGIILTGLYSENFQSVL	237		
Db	152	AFGIPVRTFTH-----EVTLWYRAPEILLGTRHYSTPVDVWSVGCIF-----AEM	197		
QY	238	VKDDKELTNDSHVSDLYLLNQIFENFGTPN-----LTDFEDELFCDEYNNENLHFK	288		
Db	198	VTQRPLFPDSEIDELF---KIFRVMTGTPNEDTWPGVVTLPDF-----	238		
QY	289	KFNLOKYPKRDWDILPRCNDLDMKEIFTKMIRYDRSKRITSKEILQ	335		
Db	238	KSAPPKWPSKDLATIVPNL-DGAGLDLDDKTSRLDPSKRITARNAL	283		

Search completed: November 6, 1999, 08:26:48
Job time: 771 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 1999, 14:48:40 ; Search time 11.2 Seconds
(without alignments)
855.621 Million cell updates/sec

Title: US-09-072-994-14
Perfect score: 1791
Sequence: 1 MKLSDYYIDKELIYNSAISD.....IRYDRSKRITSKEILQLMLD 339

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	276.5	15.4	349	1 STY1_SCHPO	Q09892 schizosacch
2	273	15.2	296	1 CC2_DICDI	P34112 dictyosteli
3	267	14.9	368	1 CAK1_YEAST	P43568 sacccharomyc
4	266	14.9	298	1 CDK2_MESAU	P48963 mesocricetu
5	265	14.8	297	1 CDK2_XENLA	P23437 xenopus lae
6	262.5	14.7	294	1 CC2_MAIZE	P23111 zea mays (m
7	262	14.6	298	1 CDK2_HUMAN	P24941 homo sapien
8	262	14.6	391	1 KC21_CHICK	P21868 gallus gall
9	262	14.6	391	1 KC21_HUMAN	P19138 homo sapien
10	262	14.6	391	1 KC21_RAT	P19139 rattus norv
11	261	14.6	298	1 CDK2_CARAU	P43450 carassius a
12	261	14.6	306	1 CSK1_SCHPO	P36615 schizosacch
13	259.5	14.5	311	1 CC23_TRYBB	P54666 trypanosoma
14	259	14.5	391	1 KC21_MOUSE	Q60737 mus musculu
15	259	14.5	350	1 KC22_XENLA	P28020 xenopus lae
16	258.5	14.4	294	1 CC21_ORYSA	P29618 oryza sativ
17	258	14.4	415	1 ERK1_DICDI	P42525 dictyosteli
18	258	14.4	391	1 KC21_RABIT	P33674 oryctolagus
19	256	14.3	305	1 CDK3_HUMAN	Q00526 homo sapien
20	255.5	14.3	297	1 CC2_SCHPO	P04551 schizosacch
21	255.5	14.3	435	1 HOG1_YEAST	P32485 sacccharomyc
22	255.5	14.3	424	1 JNK2_HUMAN	P45984 homo sapien
23	255.5	14.3	420	1 KC2A_THEPA	P28547 theileria p
24	255	14.2	288	1 CC2H_PLAFK	Q07785 plasmodium
25	254.5	14.2	302	1 CC2_CARAU	P51958 carassius a
26	254.5	14.2	423	1 JNK2_RAT	P49186 rattus norv
27	251	14.0	350	1 KC22_CHICK	P21869 gallus gall
28	250.5	14.0	302	1 CC22_XENLA	P24033 xenopus lae
29	250	14.0	350	1 KC22_BOVIN	P20427 bos taurus
30	249	13.9	377	1 HOG1_CANAL	Q92207 candida alb
31	248.5	13.9	332	1 KC21_SCHPO	P40231 schizosacch
32	248	13.8	350	1 KC22_HUMAN	P19784 homo sapien
33	248	13.8	424	1 KC47_ORYSA	P29620 oryza sativ
34	246	13.7	317	1 CC28_CANAL	P43063 candida alb
35	246	13.7	361	1 MP38_XENLA	P47812 xenopus lae
36	245.5	13.7	302	1 CC21_XENLA	P35567 xenopus lae
37	245	13.7	333	1 KC22_ARATH	Q08466 arabidopsis
38	245	13.7	337	1 KC2A_DICDI	Q02720 dictyosteli
39	244.5	13.7	291	1 CC21_MEDSA	P24923 medicago sa
40	243.5	13.6	294	1 CC2A_ARATH	P24100 arabidopsis
41	243	13.6	387	1 MMK1_MEDSA	Q07176 medicago sa
42	242.5	13.5	294	1 CC2_VIGUN	P52389 vigna ungu
43	241	13.5	361	1 MP38_CYPCA	Q90336 cyprinus ca

44 240.5 13.4 303 1 CC2_CHICK P13863 gallus gall
45 240.5 13.4 297 1 CC2_HUMAN P06493 homo sapien

ALIGNMENTS

RESULT 1
STY1_SCHPO
ID STY1_SCHPO STANDARD; PRT; 349 AA.
AC Q09892;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MITOGEN-ACTIVATED PROTEIN KINASE STY1 (EC 2.7.1.1.-) (MAP KINASE STY1).
GN STY1 OR SPC1 OR SPAC24B11.06C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA MILLAR J.B.A.; BUCK V.; WILKINSON M.G.;
RL SUBMITTED (JUN-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RA SHIOZAKI K.; RUSSELL P.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA ODELL C.; CHURCHER C.M.; BARRELL B.G.; RAJANDREAM M.A.; WALSH S.V.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: INVOLVED OSMOREGULATION AND STRESS RESPONSE PATHWAYS.
CC PROBABLY ACTIVATED BY WIS1.
CC -!- ENZYME REGULATION: ACTIVATED BY TYROSINE AND THREONINE
CC PHOSPHORYLATION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CDC2/CDC28 SUBFAMILY OF SER/THR
CC PROTEIN KINASES. STRONGEST SIMILARITY WITH OTHER MAP KINASES.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; X89262; G897810; -
EMBL; U26739; G1022685; -
EMBL; Z67757; G1061294; -
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; Q16539; 1WFC.
DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW PHOSPHORYLATION.
FT DOMAIN 20 299 PROTEIN KINASE.
FT NP_BIND 26 34 ATP (BY SIMILARITY).
FT BINDING 49 49 ATP (BY SIMILARITY).
FT ACT_SITE 141 141 BY SIMILARITY.
FT MOD_RES 171 171 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT MOD_RES 173 173 (BY SIMILARITY).
FT MOD_RES 173 173 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT MOD_RES 173 173 (BY SIMILARITY).
SQ SEQUENCE 349 AA; 40222 MW; B3385C56 CRC32;

Query Match 15.4%; Score 276.5; DB 1; Length 349;
Best Local Similarity 26.9%; Pred. No. 2.7e-10;
Matches 93; Conservative 60; Mismatches 114; Indels 79; Gaps 15;

```
QY 2 KLSDYIDKELIYNSAISDIYTAIDKFNPNLPVCLKIVDEDFSLP--PHSIHREIFILKTL 59
    ::: | | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 15 EITTRYSDLOQIGMGAFGLVCSAKDQLTGMNVAVKKIMKPFSTPVLAKRTYRELKLLKHL 74
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 60 KPHNPIIEYFNDLKI--YDDVILVTKLYRYDLSQLIETKCYCKRTRTFIYINGNLVSNQ 117
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 75 R-HENIIS-LSDIFISPFEDIYFVTELLGTDLHRL--TSR----- 112
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 118 YTLANEIEEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICY 177
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 112 -----PLETQFIQFLYQILRGLKFVHSAGVIHRDLKPSNILINEN-----CDLKIC- 159
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 178 DLKLPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEDISWLSGIILTGLYSENFQSVL 237
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 DFLARIQDPQMTGY--VSTRYYRAPEIMLTWQKYNVEVDIWSAGCIF-----AEM 207
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 238 VKDKKELTNDSHVSDLYLLNQIFENFGTPNLTDFEDELFCDEY-----NNENLHFK 288
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 208 IEGKPLFGRDHVNOFSIITEL---LGTPPMEVI--ETICSKNTLRFVQSLPQKEKVPF- 262
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 289 KFNQKYPRKQDWDIILPRCNDLDMKEIFTKMIYDRSKRITSKEIL 334
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 262 ---AEKFNADPDAI-----DLLEKMLVDFPRKRISAADAL 294
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |

RESULT 2
CC2_DICDI
ID CC2_DICDI STANDARD; PRT; 296 AA.
AC P34112;
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CELL DIVISION CONTROL PROTEIN 2 HOMOLOG (EC 2.7.1.-) (P34 PROTEIN KINASE).
DE KINASE).
GN CDCB.
OS DICTYOSTELIUM DISCOIDEUM (SLIME MOLD).
OC EUKARYOTA; DICTYOSTELIIDA; DICTYOSTELIUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92379089.
RA MICHAELIS C., WEEKS G.;
RT "Isolation and characterization of a cdc 2 cDNA from Dictyostelium discoideum.";
RL BIOCHIM. BIOPHYS. ACTA 1132:35-42(1992).
CC -!- FUNCTION: PLAYS A KEY ROLE IN THE CONTROL OF THE EUKARYOTIC CELL CYCLE. IT IS REQUIRED IN HIGHER CELLS FOR ENTRY INTO S-PHASE AND MITOSIS. P34 IS A COMPONENT OF THE KINASE COMPLEX THAT PHOSPHORYLATES THE REPETITIVE CARBOXYL-TERMINUS OF RNA POLYMERASE II.
CC -!- ENZYME REGULATION: PHOSPHORYLATION AT THR-20 OR TYR-21 INACTIVATES THE ENZYME, WHILE PHOSPHORYLATION AT THR-162 ACTIVATES IT (BY SIMILARITY).
CC -!- SUBUNIT: FORMS A STABLE BUT NON-COVALENT COMPLEX WITH A REGULATORY SUBUNIT AND WITH A CYCLIN.
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-PROTEIN KINASES. BELONGS TO THE CDC2/CDKX SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
CC EMBL; M80808; G167686; -.
DR PIR; S24386; S24386.
DR DICTYDB; DD05040; CDCB.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P24941; 1A01.
```

```
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW CELL CYCLE; CELL DIVISION; MITOSIS; PHOSPHORYLATION.
FT DOMAIN 10 288
FT NP_BIND 16 24
FT BINDING 39 39
FT ACT_SITE 129 129
FT MOD_RES 20 20
FT MOD_RES 21 21
FT MOD_RES 162 162
SQ SEQUENCE 296 AA; 33773 MW; 6AABD40A CRC32;
    15.2%; Score 273; DB 1; Length 296;
    Best Local Similarity 26.0%; Pred. No. 3.6e-10;
    Matches 86; Conservative 63; Mismatches 122; Indels 60; Gaps 13;

QY 7 YIDKELIYNSAISDIYTAIDKFNPNLPVCLKIVD-EDFSLPHSIHREIFILKTLKPHPN 65
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 YQKLEKLGEGTYGKVKAKEKATGRMVALKKIRLEDDGVPTAL-REISLLKEV-PHPNV 67
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 66 IEYFNDLKIYDDVILVTKLYRYDLSQLIE-ITKYCKRTRTFIYINGNLVSNQYTLANEI 124
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 68 VSLFDVLHCQNRLLYVEYLDQDLKKYSDVSPALCPQL----- 106
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 125 EEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFICYDLKLP 184
    | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 106 ----IKSYLYQLLKGLAYSHGHRILHRDLKPNLLIDRQALK--LADFGLARAVSIPVR 159
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 185 DEPPMAKYIDVSTGIYKAPELILGITNYEYEDISWLSGIILTGLYSENFQSVLVKDDKEL 244
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 160 ----VYTHEIVTLWYRAPEVLGSKSYSPVDMWSVGC-----FGMLNK--KPL 204
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 245 -TNDSHVSDLYLLNQIFENFGTPNLTDFEDELFCDEYNNENLHFKFNQKYPRKQDWDII 303
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 205 FSGDCEIDQIF--RIFRVLGTPDDSIWPGVTKLPEYVS-----TFPNWPGQPYNKI 253
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |
QY 304 LPRCNDLDMKEIFTKMIYDRSKRITSKEIL 334
    | | | | : : | | : : | | : : | | : : | | : : | | : : | |
Db 254 FPRC-EPLALDLIAKMLQYEPSKRISAKEAL 283
    : | | | | : : | | : : | | : : | | : : | | : : | | : : | |

RESULT 3
CAK1_YEAST
ID CAK1_YEAST STANDARD; PRT; 368 AA.
AC P43568;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SERINE/THREONINE-PROTEIN KINASE CAK1 (EC 2.7.1.-) (CDK-ACTIVATING KINASE).
DE CAK1 OR CIV1 OR YFL029C.
GN CAK1 OR CIV1 OR YFL029C.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
OC SACCHAROMYCETACEAE; SACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96361353.
RA KALDIS P., SUTTON A., SOLOMON M.J.;
RT "The Cdk-activating kinase (CAK) from budding yeast.";
RL CELL 86:553-564(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE; 95400292.
RA MURAKAMI Y., NAITOU M., HAGIWARA H., SHIBATA T., OZAWA M.,
RA SASANUMA S.-I., SASANUMA M., TSUCHIYA Y., SOEDA E., YOKOYAMA K.,
RA YAMAZAKI M., TASHIRO H., EKI T.;
RT "Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae.";
RL NAT. GENET. 10:261-268(1995).
CC -!- SIMILARITY: BELONGS TO THE CDC2/CDC28 SUBFAMILY OF SER/THR-PROTEIN KINASES.
-----
CC
```



```
DR EMBL; M68520; G180178; -.
DR PIR; A41227; A41227.
DR PIR; S16520; S16520.
DR PIR; S17873; S17873.
DR PDB; 1FIN; 27-JAN-97.
DR PDB; 1HCK; 07-DEC-96.
DR PDB; 1HCL; 07-DEC-96.
DR PDB; 1AQL; 12-NOV-97.
DR PDB; 1JST; 11-JAN-97.
DR PDB; 1JSU; 29-JUL-97.
DR PDB; 1BUH; 09-SEP-98.
DR PDB; 1B38; 23-DEC-98.
DR PDB; 1B39; 23-DEC-98.
DR MIM; 116953; -.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW CELL CYCLE; CELL DIVISION; MITOSIS; PHOSPHORYLATION; 3D-STRUCTURE.
FT DOMAIN 4 286
FT NP_BIND 10 18 ATP (BY SIMILARITY).
FT BINDING 33 33 ATP (BY SIMILARITY).
FT ACT_SITE 127 127 BY SIMILARITY.
FT MOD_RES 14 14 PHOSPHORYLATION.
FT MOD_RES 15 15 PHOSPHORYLATION.
FT MOD_RES 160 160 PHOSPHORYLATION (BY CAK).
FT MUTAGEN 14 14 T->A: INCREASE ACTIVITY 2 FOLD.
FT MUTAGEN 15 15 Y->F: INCREASE ACTIVITY 2 FOLD.
FT MUTAGEN 160 160 T->A: ABOLISHES ACTIVITY.
SQ SEQUENCE 298 AA; 33929 MW; 7538D6FA CRC32;

Query Match 14.6%; Score 262; DB 1; Length 298;
Best Local Similarity 26.2%; Pred. No. 1.7e-09;
Matches 83; Conservative 59; Mismatches 119; Indels 56; Gaps 11;

QY 21 IYTAIDKFNLPVCLKIV--DEDESLPPHSIHREIFILKTLKPHNPNIIEYFNLDKIYDDV 78
Db 18 VYKARNKLTGEVVALKKIRLDTETEGVPSTAIRESLLKELN-HPNIVKLLDVIHTENKL 76
QY 79 ILVTKLYRYDLSQLIEITKYCKRTTRFIYGINGNLVSNQYTLANEIEEKDIKLWLKSMSS 138
Db 77 YLVFEFLHQDLKKFMDASA-----LTGIPLPL-----IKSYLFQLLQ 113
QY 139 GLEFIHSQGIHRDIKPSNIFFAADDITQPIIGDFDICYDLKLPKDEPPMAKYI-DVST 197
Db 114 GLAFCHSHRVLHRLDKPQNLLINTEGAIK--LADFLARAFGV-----PVRTYTHEVVT 165
QY 198 GIYKAPELIGITNYEYIDISLGIILTGLYSENFQSVLVKDKKELTNDSHVSDIYLLN 257
Db 166 LWYRAPEILGCKYYSTAVDIWSLGCIF-----AEMVTRRALFPDSEIDQLF--- 214
QY 258 QIFENFGTNPNTDFEDELFCDEYNENLHFKKFNKQYPRKQWDIILPRCNDLDMKEIFT 317
Db 214 RIFRTLGTDPDEVVWPGVTSMPDY-----KPSFPKWQARQDFSKVVPPLDED-GRSLLS 264
QY 318 KMIRYDRSKRITSKEIL 334
Db 265 QMLHYDPNKRISAKAAL 281

RESULT 8
KC21_CHICK STANDARD; PRT; 391 AA.
AC P21868;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE CASEIN KINASE II, ALPHA CHAIN (CK II) (EC 2.7.1.37).
OS GALLUS GALLUS (CHICKEN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
```

```
RN [1] SEQUENCE FROM N.A.
RP MEDLINE; 91115855.
RX MARIDOR G., PARK W., KREK W., NIGG E.A.;
RT "Casein kinase II. cDNA sequences, developmental expression, and
RT tissue distribution of mRNAs for alpha, alpha', and beta subunits of
RT the chicken enzyme.";
RL J. BIOL. CHEM. 266:2362-2368(1991).
CC -!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR
CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS
CC AS SUBSTRATES.
CC -!- FUNCTION: THE ALPHA AND ALPHA' CHAINS CONTAIN THE CATALYTIC SITE.
CC -!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA CHAIN, AN ALPHA' AND TWO
CC BETA CHAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59456; G211534; -.
DR PIR; A38611; A38611.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P28523; 1A60.
KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.
FT DOMAIN 39 324 PROTEIN KINASE.
FT NP_BIND 45 53 ATP (BY SIMILARITY).
FT BINDING 68 68 ATP (BY SIMILARITY).
FT ACT_SITE 156 156 BY SIMILARITY.
SQ SEQUENCE 391 AA; 45190 MW; EDB98195 CRC32;

Query Match 14.6%; Score 262; DB 1; Length 391;
Best Local Similarity 26.5%; Pred. No. 2.3e-09;
Matches 90; Conservative 59; Mismatches 125; Indels 66; Gaps 13;

QY 5 DYYIDKELIYNSAISDIYTAIDKFNLPVCLKIVDEDFSLPHSIHREIFILKTLKPHNP 64
Db 38 DYQLVRKL-GRGKYSEVFEAINITNNEKVVVKILK--PVKKKKIKKILENLRGGPN 93
QY 65 IIEYFNLDKIYDDVILVTKLYRYDLSQLIEITKYCKRTTRFIYGINGNLVSNQYTLANEI 124
Db 94 IITLADIVK-DPV-----SRTPALVFEHVNTDFKQ--LYQTL 128
QY 125 EEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFAADDITQPIIGDFDICYDLKLPK 184
Db 129 TDYDIRFYKVEILKALDYCHSMGIMHRDVKPHNMIDHEHRKRLI-----DWGLAEF 181
QY 185 DEPPMAKYIDVSTGIYKAPELILGITNYEYIDISLGIILTGLY--SENFQSVLVKDDK 242
Db 182 YHPGQYNNVRVASRYFKGPELLVDYQMYDYSLDMSWLSGCLASMIFFKEPF----- 233
QY 243 ELTNDSHVSDLY-LLNQIFENFGTNPNTDFEDELFCDEYNENLHFKKFN--LQKYPRKD 299
Db 233 -----FHGHDNYDQLVRIAKVLGTEDLYD-----YIDKYNIE--LDPRENDILGRHSKR 280
QY 300 WDIILPRCNDLDMK---EIFTKMIRYDRSKRITSKEILQ 335
Db 281 WERFVHSENQHLVSPALDPLDKLLRYDHQSRLTAREAME 320

RESULT 9
KC21_HUMAN
ID KC21_HUMAN STANDARD; PRT; 391 AA.
AC P19138; P20426;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
```



```
GN CSK1 OR SPAC1D4.06C.
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;
OC SCHIZOSACCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 93223713.
RA MOLZ L., BEACH D.;
RT "Characterization of the fission yeast mcs2 cyclin and its associated
  protein kinase activity.";
RL EMBO J. 12:1723-1732(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA LYE G., CHURCHER C.M., BARRELL B.G., RAJANDREAM M.A., WALSH S.V.;
RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -!- FUNCTION: NOT KNOWN. COULD FUNCTION IN ASSOCIATION WITH CYCLIN
  MCS2.
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-
  PROTEIN KINASES. BELONGS TO THE CDC2/CDKX SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S59896; G299549; -.
DR EMBL; Z69239; E1188809; -.
DR PIR; S39151; S39151.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P24941; 1AQ1.
DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.
KW DOMAIN 11 306 PROTEIN KINASE.
FT NP_BIND 17 25 ATP (BY SIMILARITY).
FT BINDING 40 40 ATP (BY SIMILARITY).
FT ACT_SITE 129 129 BY SIMILARITY.
SQ SEQUENCE 306 AA; 34672 MW; 514DE9CE CRC32;

Query Match 14.6%; Score 261; DB 1; Length 306;
Best Local Similarity 25.1%; Pred. No. 2e-09;
Matches 88; Conservative 62; Mismatches 127; Indels 74; Gaps 15;

QY 1 MKLSDDYI---DKELIYNSAISDIYTAIDKFNNLPVCLKIVDEDFSLPPHSIHREIFIL 56
Db 1 MKSVGHFVPLTDIRHLTDGTISEVFVGERKNSKKLYVIKVQGLVFKRPPHDAMRGKLIL 60

QY 57 KTL-KPH-PNIIIEVFNDLKIYDDVILVTKLYRYDLSQL---IEITKYCKRTRTFIYGING 111
Db 61 ESIGHPHIERIVDSFIDNEA-GSVYLITSPKSFVLSVDMDEISIDTKCKIVLQ----- 113

QY 112 NLVSNQYTLANEIEEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIPFARDDDITQP-II 170
Db 113 -----ISSALEYLEKHGILHRDHPNILL--DSMNGPAYL 146

QY 171 GDFDICYDLKLPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEIDWSLGIILTGLYS 230
Db 147 SDFSIAWSKQHPGEEVQELIP--QIGTGHYRAIETLFGCHSYGHEVDVRTWTFGILIAELFS 204

QY 231 ENFQSVLVKDKELTNDSHVSDLYLLNQIFENFGTPN-----LTDFEDELFCDEYNNEN 284
Db 205 NQ-----ALFDDG--SSEGWPSELRLTASSIIQTGTNPNSMWPBELSTFPD----- 248

QY 285 LHFKKFNLOQYPRKDWDIILPRCNDLDMKEIFTKMIYDRSKRITSKEILQ 335
Db 248 --WNKEFIFHEYPKPWSEILPSV-DTSIQYIVSHLVTY--SNRASPSFVIE 293
```

```
RESULT 13
CC23_TRYBB
ID CC23_TRYBB STANDARD; PRT; 311 AA.
AC P54666;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE CELL DIVISION CONTROL PROTEIN 2 HOMOLOG 3 (EC 2.7.1.-).
GN CRK3.
OS TRYPANOSOMA BRUCEI BRUCEI.
OC EUKARYOTA; EUGLENOZOA; KINETOPLASTIDA; TRYPANOSOMATIDAE; TRYPANOSOMA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISTAT;
RX MEDLINE; 96009893.
RA MOTTRAM J., SMITH G.;
RT "A family of trypanosome cdc2-related protein kinases.";
RL GENE 162:147-152(1995).
CC -!- FUNCTION: PROBABLY INVOLVED IN THE CONTROL OF THE CELL CYCLE.
CC -!- ENZYME REGULATION: PHOSPHORYLATION AT SER-33 OR TYR-34 INACTIVATES
  THE ENZYME.
CC -!- SUBUNIT: FORMS A STABLE BUT NON-COVALENT COMPLEX WITH A REGULATORY
  SUBUNIT AND WITH A CYCLIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CDC2/CDC28 SUBFAMILY OF SER/THR
  PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X74617; G397365; -.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00069; pkinase; 1.
DR HSSP; P24941; 1AQ1.
DR TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING;
KW PHOSPHORYLATION.
FT DOMAIN 23 306 PROTEIN KINASE.
FT NP_BIND 29 37 ATP (BY SIMILARITY).
FT BINDING 52 52 ATP (BY SIMILARITY).
FT ACT_SITE 145 145 BY SIMILARITY.
FT MOD_RES 33 33 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 34 34 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 311 AA; 35047 MW; 6594D66E CRC32;

Query Match 14.5%; Score 259.5; DB 1; Length 311;
Best Local Similarity 25.9%; Pred. No. 2.4e-09;
Matches 91; Conservative 57; Mismatches 113; Indels 91; Gaps 15;

QY 5 DYYIDKELIYNSAISDIYTAIDKFNNLPVCLKIV-----DEDFSLPPHSIHREIFILKTL 59
Db 21 DRYNRMDILGEGYGVVYRAVDRATGQIVALKKVRDLRTDEGI---PQTALREVSIHQEI 77

QY 60 KPHPNIIIEYFN---DLKIYDDVILVTKLYRYDLSQLIEITKYCKRTRTFIYGINGNLVS 115
Db 78 H-HPNIVNLLDVICADGKLY---LIFEYVDHDLKKALE-----KRGGAFA----- 118

QY 116 NQYTLANEIEEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIPFARDDDITQPIIGDFDI 175
Db 118 -----TGTTLKKIYQLLEGSLFCHRRHVRDLKPANILVTTDNSVK--IADFGL 166

QY 176 CYDLKLPKDEPPMAKYI-DVSTGIYKAPELILGITNYEYEIDWSLGIILTGLYSENFQ 234
Db 167 ARAFQI-----PMHTYTHEVVTILWYRAPEILLGEKHHTPAVDWMWSIGCIFAEELAR---G 217
```

QY	235	SVLVKDDKELTNDSHVSDLYLLNQIFENFGTP-----NLTFDEELFCDEYNNNE	283
Db	218	KVLFRGDSEIGQ-----LFEIFQVLGTPMDAEGSWLGVSSLPDYRDV-----	260
QY	284	NLHFKKFNLOQYPRKRDWDIILPRCNDLDMKEIFTKMIRYDRSKRITSKEILQ	335
Db	260	-----FPKWSGRPLTQVLPRTLGDGDAV-DLLSQMLRYNPAERISAKAALQ	302
RESULT 14			
ID	KC21_MOUSE	STANDARD;	PRT; 391 AA.
AC	Q60737;		
DT	15-JUL-1998	(REL. 36, CREATED)	
DT	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)	
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)	
DE	CASEIN KINASE II, ALPHA CHAIN (CK II) (EC 2.7.1.37);		
GN	CSNK2A1 OR CKIIA.		
OS	MUS MUSCULUS (MOUSE).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FVB/N; TISSUE=SPLEEN;		
RX	MEDLINE; 95149131.		
RA	SELDIN D.C., LEDER P.;		
RT	"Casein kinase II alpha transgene-induced murine lymphoma: relation to theileriosis in cattle.";		
RL	SCIENCE 267:894-897(1995).		
CC	-!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS AS SUBSTRATES.		
CC	-!- FUNCTION: THE ALPHA AND ALPHA' CHAINS CONTAIN THE CATALYTIC SITE.		
CC	-!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA CHAIN, AN ALPHA' AND TWO BETA CHAINS.		

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U17112; G665543; -.		
DR	MGD; MGI:88543; CSNK2A1.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.		
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
DR	PFAM; PF00069; pkinase; 1.		
DR	HSSP; P28523; 1A60.		
DR	TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.		
KW	DOMAIN 39 324		
FT	NP_BIND 45 53		
FT	BINDING 68 68		
FT	ACT_SITE 156 156		
FT	SEQUENCE 391 AA; 45161 MW; 91C347DB CRC32;		
Query Match			
Best Local Similarity 14.5%; Score 259; DB 1; Length 391;			
Matches 90; Conservative 58; Mismatches 126; Indels 66; Gaps 13;			
QY	5	DYIDKELIYNSAISIYTAIDKFNLPVCLKIVDEDFSLPPHSIHREIFILKTLKPHN	64
Db	38	DYQLVRKL-GRGKYSEVFEAINITNNEKVVVKILK---PVKKKKIKREIKILENLRGPN	93
QY	65	IIIEYFNDLKIYDDVILVTKLRYDLSQLJEITKYCKRTTRFTYINGNLSNQYTLANEI	124
Db	94	IITLADIVK--DPV-----SRTPALVFEHVNTDFKQ--LYQTL	128
QY	125	EEDKIKLWLKMSGGLFIHSQGIHHRDIKPSNIFFARDDDITPIGDFDICYDLKLPK	184

Db	129	TDYDIRFYMEILKALDYCHSMGIMHRDVKPHNVMIDHEHRKRLI-----DWGLAEF	181
QY	185	DEPPMAKYIDVSTGIYKAPELILGITNVEYEIDISLGIILTGLY--SENFQSVLVKDDK	242
Db	182	YHPGOEYNVRVASRYFKGPELLVDYQMYDYSLDMSLGCMLASMIKKEPF-----	233
QY	243	ELTNDSHVSDLY-LLNQIFENFGTPNLTDFEDELFCDEYNNENLHFKKFN--LQKYPRKD	299
Db	233	-----FHGHDNYDQLVRIAKVLGTEDLYD----YIDKYNIE--LDPRFNDILGRHSKR	280
QY	300	WDIILPRCNDLDMK----EIFTKMIRYDRSKRITSKEILQ	335
Db	281	WERFVHSENOHLVSPALDFLDKLLRYDHQSRLTVREAME	320
RESULT 15			
ID	KC22_XENLA	STANDARD;	PRT; 350 AA.
AC	P28020;		
DT	01-AUG-1992	(REL. 23, CREATED)	
DT	01-AUG-1992	(REL. 23, LAST SEQUENCE UPDATE)	
DT	01-NOV-1995	(REL. 32, LAST ANNOTATION UPDATE)	
DE	CASEIN KINASE II, ALPHA' CHAIN (CK II) (EC 2.7.1.37).		
CS	XENOPUS LAEVIS (AFRICAN CLAWED FROG).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; AMPHIBIA; BATRACHIA; ANURA;		
OC	MESOBATRACHIA; PIPOIDEA; PIPIDAE; XENOPODINAE; XENOPUS.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=OVARY;		
RX	MEDLINE; 92183811.		
RA	JEDLICKI A., HINRICHS M.V., ALLENDE C., ALLENDE J.E.;		
RT	"The cDNAs coding for the alpha- and beta-subunits of Xenopus laevis casein kinase II.";		
RL	FEBS LETT. 297:280-284(1992).		
CC	-!- FUNCTION: CASEIN KINASES ARE OPERATIONALLY DEFINED BY THEIR PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS AS SUBSTRATES.		
CC	-!- FUNCTION: THE ALPHA AND ALPHA' CHAINS CONTAIN THE CATALYTIC SITE.		
CC	-!- SUBUNIT: TETRAMER COMPOSED OF AN ALPHA CHAIN, AN ALPHA' AND TWO BETA CHAINS.		

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X62375; G64628; -.		
DR	PIR; S18897; S18897.		
DR	PIR; S20404; S20404.		
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.		
DR	PROSITE; PS00108; PROTEIN_KINASE_ST; 1.		
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.		
DR	PFAM; PF00069; pkinase; 1.		
DR	HSSP; P28523; 1A60.		
KW	TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING.		
FT	DOMAIN 39 324		
FT	NP_BIND 45 53		
FT	BINDING 68 68		
FT	ACT_SITE 156 156		
FT	SEQUENCE 350 AA; 41454 MW; 8956ED02 CRC32;		
Query Match			
Best Local Similarity 14.5%; Score 259; DB 1; Length 350;			
Matches 90; Conservative 58; Mismatches 125; Indels 66; Gaps 13;			
QY	5	DYIDKELIYNSAISIYTAIDKFNLPVCLKIVDEDFSLPPHSIHREIFILKTLKPHN	64
Db	38	DYQLVRKL-GRGKYSEVFEAINITNNEKVVVKILK---PVKKKKIKREIKILENLRGPN	93

QY	65	IIEYFNDLKIYDDVILVTKLYRYDLSQLIEITKYCKRTRFRIFYGINGNLVSQYTLANEI	124
Dd	94	IITLADIVK--DPV-----SRPALVFHVNTDFKQ--LYQTL	128
QY	125	EEDIKLWLKSMSSGLEFIHSQGIIRHDIKPSNIFFARDITQPIIGDFDICVDKLPPK	184
Dd	129	TDYDIRFYMEILKALDYCHSMGMHRDVKPHNVMDHEHKKRLRI-----DWGLAEF	181
QY	185	DEPMAKYIDVSTGIYKAPELILGITNYEYEIDIWSLGILTGLY--SENFQSVLVKDDK	242
Dd	182	YHPGQEYNVRVASRYFKGPPELLVDYQMYDYSLDWMSLCGLASMIFRKEPF-----	233
QY	243	ELTNDSHVSDLY-LLNQIFENFGTPNLTFDELEFCDEYNNENLHFKKFN--LQKYPRKD	299
Dd	233	----FHGHNDYDLVRIAKVLGTEDLYD----YIDKYNIE--LDPRFNDILGRHSKR	280
QY	300	WDIILPCRNDDLKMK----EIFTKMIRYDRSKRITSKEIL	334
Dd	281	WERFVHSENQHVLVSPEALDFDLKLLRYDHQTRLTAREAM	319

Search completed: November 6, 1999, 00:45:07
Job time: 35787 sec


```
QY 2 KLSDDYIDKELIYNSAIDYTAIDKFNLPVCLKIVDEDF--SLPPHSIHREIFILKTL 59
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 20 ELPKTYVSPTHVSGAGYSCSAIDKRSGEKVAIKKLSRPFQSEIFAIRAYRELLLLKHM 79
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 60 KPHPNII---EYF---NDLKIIDVILVTKLYRYDLSQLIEITKYCKRTRFIYGINL 113
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 80 Q-HENVIGLLDVFTPASSLGNFYDFYLVMPFMTDLQKI----- 118
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 114 VSNQYTLANEIEEKDIKWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDF 173
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 118 -----MGMEFSEEKIQYLVYQMLKGLKYIHSAGVVHRDLKPGNL-AVNEDCELKIL--- 168
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 174 DICYDLKLPPKDEPPMAKYIDVSTGIYKAPELILGITNYEYIDISLGLI-----LTGLY 229
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 168 ----DLGLARHADAEMTGY--VVTRWYRAPEVILSWMHYNQTVDIWSVGCIMAEMLTG-- 220
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 230 SENFQSVLVKDDKELTNDSHVSDLYLLNQIFENFGTGNLTDFEDELFCDEYNENLHFKK 289
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 220 -----KTLFKGKDYLDQ-----LTQILKVTGVPG-TEFVQKL-----NDKAAKSYI 259
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 290 FNLQYPRKDWDIILPRCNDLDMKEIFTKMIRYDRSKRITSKEIL 334
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 260 QSLPQTPRKDFTQLFPRASPQ-PADLLEKMLELDVDKRLTAAQAL 303
:| | | | | : | | | | | : | | | | | : | | | | | :
```

```
RESULT 6
US-08-920-296-2
: Sequence 2, Application US/08920296
: Patent No. 5846778
: GENERAL INFORMATION:
: APPLICANT: Hawkins, Phillip R.
: APPLICANT: Au-Young, Janice
: APPLICANT: Guegler, Karl J.
: APPLICANT: Wilde, Craig G.
: TITLE OF INVENTION: A NOVEL HUMAN MAP KINASE HOMOLOG
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/920,296
: FILING DATE: 28-AUG-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/674,612
: FILING DATE:
: APPLICATION NUMBER: 60/000,722
: FILING DATE: 30-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0036 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 365 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: IMMEDIATE SOURCE:
: LIBRARY: Stomach
```

```
; CLONE: 214915
US-08-920-296-2
Query Match 14.0%; Score 251.5; DB 2; Length 365;
Best Local Similarity 26.4%; Pred. No. 4.1e-15;
Matches 91; Conservative 56; Mismatches 125; Indels 73; Gaps 14;
QY 2 KLSDDYIDKELIYNSAIDYTAIDKFNLPVCLKIVDEDF--SLPPHSIHREIFILKTL 59
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 20 ELPKTYVSPTHVSGAGYSCSAIDKRSGEKVAIKKLSRPFQSEIFAIRAYRELLLLKHM 79
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 60 KPHPNII---EYF---NDLKIIDVILVTKLYRYDLSQLIEITKYCKRTRFIYGINL 113
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 80 Q-HENVIGLLDVFTPASSLGNFYDFYLVMPFMTDLQKI----- 118
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 114 VSNQYTLANEIEEKDIKWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDF 173
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 118 -----MGMEFSEEKIQYLVYQMLKGLKYIHSAGVVHRDLKPGNL-AVNEDCELKIL--- 168
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 174 DICYDLKLPPKDEPPMAKYIDVSTGIYKAPELILGITNYEYIDISLGLI-----LTGLY 229
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 168 ----DLGLARHADAEMTGY--VVTRWYRAPEVILSWMHYNQTVDIWSVGCIMAEMLTG-- 220
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 230 SENFQSVLVKDDKELTNDSHVSDLYLLNQIFENFGTGNLTDFEDELFCDEYNENLHFKK 289
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 220 -----KTLFKGKDYLDQ-----LTQILKVTGVPG-TEFVQKL-----NDKAAKSYI 259
:| | | | | : | | | | | : | | | | | : | | | | | :
QY 290 FNLQYPRKDWDIILPRCNDLDMKEIFTKMIRYDRSKRITSKEIL 334
:| | | | | : | | | | | : | | | | | : | | | | | :
Db 260 QSLPQTPRKDFTQLFPRASPQ-PADLLEKMLELDVDKRLTAAQAL 303
:| | | | | : | | | | | : | | | | | : | | | | | :
```

```
RESULT 7
US-08-746-788-2
: Sequence 2, Application US/08746788
: Patent No. 5869043
: GENERAL INFORMATION:
: APPLICANT: McDonnell, Peter
: APPLICANT: Young, Peter
: TITLE OF INVENTION: DRUG BINDING PROTEIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: SmithKline Beecham Corporation
: STREET: 709 Swedeland Road
: CITY: King of Prussia
: STATE: PA
: COUNTRY: USA
: ZIP: 19406-0939
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq Version 1.5
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/746,788
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/468,902
: FILING DATE: 06-JUN-1995
: APPLICATION NUMBER: 08/123,175
: FILING DATE: 17-SEP-1993
: APPLICATION NUMBER: 08/250,975
: FILING DATE: 31-MAY-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Schreck, Patricia A
: REGISTRATION NUMBER: 33,777
: REFERENCE/DOCKET NUMBER: ATG50036
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-270-5031
: TELEFAX: 610-270-5090
: TELEX:
```



```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 365 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: peptide
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE: N-terminal
;   ORIGINAL SOURCE:
US-08-746-788-2

```

Query Match	13.8%;	Score 246.5;	DB 2;	Length 365;
Best Local Similarity	26.1%;	Pred. No. 1.2e-14;		
Matches 90;	Conservative 56;	Mismatches 126;	Indels 73;	Gaps 14;

QY	2	KLSDYYIDKELIYN	SAISDIYTAIDKFN	NLPVCLKIVDEDF--SLPPHSIHREIFILKTL	59	
Db	20	ELPKTYVSPTHVGS	GAYGWCSAIDKRS	GEKVAIKKLSRPQSEIFA	KRAYRELLLLKHM 79	
QY	60	KPHNP	II--EYF--	NDLKIYDDVILVT	KLYRYDLSQLIETKYCKRTRFIYINGNL 113	
Db	80	Q-HENVIGLLDV	FTPASLRNFYD	FLVMPFMQTDLQI	----- 118	
QY	114	VSNQYTLANEIEE	KDIKLWLKSMSSG	LEFIHSQGIHRDIKPSN	IFFARDDITQPIIGDF 173	
Db	118	-----MGMEF	SEEKIQYLVYQML	KGLKYIHSAGVVHRDL	KPGNL-AVNEDCELKIL--- 168	
QY	174	DICYDLKLPPKDE	PPMAKYIDVSTGI	YKAPELILGITNVEYE	IDIWSLGI---LTGLY 229	
Db	168	-----DFGLAR	HADAEMTGY--	VVTRWYRAPEVILS	MWYHYNQTVDIWSVGC	IMAEMLTG-- 220
QY	230	SENFQSVLVKDK	ELTNDSHVSDLYL	LLNQIFENFGTPNLT	DFEDELFCDEYNNENL	HFKK 289
Db	220	-----KTLFKG	KDYLDQ-----	LTQILKVTGVPG-	TEFVQKL-----	NDKAASYI 259
QY	290	FNLOKYPRKWDI	ILPRCNDL	LMKEIFTKMIRYDRS	KRITSKEIL 334	
Db	260	QSLPQTPRKDF	TQLFPRAS	PQ-AADLLEKMLELD	VDKRLTA	AAOAL 303

```

RESULT      8
US-08-463-090B-9
; Sequence 9, Application US/08463090B
; Patent No. 5801015
; GENERAL INFORMATION:
; APPLICANT: Cottarel, Guillaume
; APPLICANT: Damagnez, Veronique
; APPLICANT: Draetta, Guilio
; TITLE OF INVENTION: Cell-Cycle Regulatory Proteins from
; TITLE OF INVENTION: Human Pathogens, and Uses Related Thereto
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley, Hoag & Elliot, LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/463,090B
; FILING DATE: 05-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709

```

```

; REFERENCE/DOCKET NUMBER: MIV032.01
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: (617) 832-1299
;
; TELEFAX: (617) 832-7000
;
; INFORMATION FOR SEQ ID NO: 9:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 317 amino acids
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: protein
;
US-08-463-090B-9

```

Query Match	13.7%;	Score	246;	DB	2;	Length	317;
Best Local Similarity	25.8%;	Pred.	No.	1.1e-14;			
Matches	88;	Conservative	69;	Mismatches	122;	Indels	62;
Gaps	15;						
QY	1	MKLSDDYIDKELIYNSAISDIYTAID-KFNPLPCLK----	IVDEDFSLPPHSIHREIFIL	56			
Db	2	VELSDYQ-RQEKVGEGTYGVYKALDTKHNNRVVALKKIRLESEDEGVPSTAI-REISLL	59				
QY	57	KTLPKHPNIIIEYFNDLKIYDDVILVT-KLYRYDLSQLIEITKYCKRTRTFIYINGNLVS	115				
Db	60	KEMKDD-----NIVRLYDIHSDSHKLYLVFEFLDLCLKYMESIPQGV-GLGANM--	110				
QY	116	NOYTLANEIEEKDIKLWLKSMSSGLEFIHSQGIHHRDIKPSNIFFARDDITQPIIGDFDI	175				
Db	110	-----IKRFMNQLIRGIKHCCHSHRVLHRDLKPQNLLIDKEGNLK--LADFG	154				
QY	176	CYDLKLPPKDEPPMAKYIDVSTGIYKAPELILGITNVEYEIDIWSLGIILTGLYSENFQS	235				
Db	155	ARAFGVPLR-----AYTHEVVTWLYRAPEILLGGKYSTGVDMWSVGCIFAEKMCNR----	206				
QY	236	VLVKDDKEL-TNDSHVSDLYLLNQIFENFGTPNLTDPEDELFCDEYNNENLHFKKFNL-Q	293				
Db	206	-----KPLFPGDSEIDEIF--RIFRILGTPNEEIWPDVNYLPDFKSSFQQWKKKPLSE	256				
QY	294	KYPRKDWDIILPRCNDLDMKEIFTKMIRYDRSKRITSKEIL	334				
Db	257	AVPSLDANGI-----DLLDQMLVYDPSRRISAKRAL	287				

RESULT 9
 US-08-874-347-18
 ; Sequence 18, Application US/08874347
 ; Patent No. 5863741
 ; GENERAL INFORMATION:
 ; APPLICANT: Limper, Andrew H.
 ; APPLICANT: Leof, Edward B.
 ; APPLICANT: Thomas, Charles F.
 ; APPLICANT: Gustafson, Michael P.
 ; TITLE OF INVENTION: CDC2 PROTEIN KINASE FROM PNEUMOCYSTIS
 ; TITLE OF INVENTION: CARINII
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Richardson P.C., P.A.
 ; STREET: 60 South Sixth Street, Suite 3300
 ; CITY: Minneapolis
 ; STATE: MN
 ; COUNTRY: USA
 ; ZIP: 55402
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/874,347
 ; FILING DATE: 13-JUN-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Ellinger, Mark S.
 ; REGISTRATION NUMBER: 34,812


```
;
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellinger, Mark S.
; REGISTRATION NUMBER: 34,812
; REFERENCE/DOCKET NUMBER: 07039/055001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-335-5070
; TELEFAX: 612-288-9696
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-874-347-23
```

```
Query Match 13.3%; Score 238.5; DB 2; Length 297;
Best Local Similarity 25.1%; Pred. No. 4.6e-14;
Matches 84; Conservative 64; Mismatches 126; Indels 61; Gaps 14;

QY 7 YIDKELIYNSAIDIIYTAIDKFNLPVCLK---IVDEDFSLPPHSIHREIFILKTLKPHP 63
Db 4 YIKIEKIGEGTYGVVYKGRHRTTGQIVAMKKIRLESEEGVPSTAI-REISLLKELR-HP 61

QY 64 NIIYFNLDKIYDD-VILVTKLYRYDLSQLEIETKYCKRTTRFIYGINGNLVSNQYTLAN 122
Db 62 NIVS-LQVLMQDSRLYLIFEFLSMDLKKYLD-----SIPPGQF---- 100

QY 123 EIEEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIIFFARDDITQPIIGDFDICYDLKLP 182
Db 100 -MDSSLVKSILYQILQGIIVFCHSRVLRDLKPQNLLI--DDKGTIKLADFLGLARAFGI- 156

QY 183 PKDEPPMAKYI-DVSTGIYKAPELILGITNYEYEDISLGIILTGLYSENFQSVLVKDD 241
Db 156 -----PIRVYTHEVTLWYRSPEVLLGSARYSTPVDIWSIGTIFAELATK----- 201

QY 242 KELTN-DSHVSDLYLLNQIFENFGTPNLTDFEDELFCDEYNENLHFKKFNQKYPRKDW 300
Db 201 KPLFHGDSIDQLF---RIFRALGTPNNEVWPEVESLQDYKNTFPKWKPGSLASHVKN-- 256

QY 301 DIILPRCNDLDMKEIFTKMIRYDRSKRITSKEILQ 335
Db 256 -----LDENGLDLLSKMLVYDPAKRISGMALK 283
```

```
RESULT 12
US-08-176-620A-16
; Sequence 16, Application US/08176620A
; Patent No. 5595904
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G.
; APPLICANT: Cobb, Melanie H.
; APPLICANT: Yancopoulos, George D.
; APPLICANT: Nye, Steven
; APPLICANT: Panayotatos, Nikos
; TITLE OF INVENTION: A Family of Map2 Protein Kinases
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

```
;
; APPLICATION NUMBER: US/08/176,620A
; FILING DATE: 03-JAN-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6526-123
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 297 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-176-620A-16

Query Match 13.3%; Score 237.5; DB 1; Length 297;
Best Local Similarity 25.4%; Pred. No. 5.6e-14;
Matches 85; Conservative 60; Mismatches 128; Indels 61; Gaps 14;

QY 7 YIDKELIYNSAIDIIYTAIDKFNLPVCLK---IVDEDFSLPPHSIHREIFILKTLKPHP 63
Db 4 YTKIEKIGEGTYGVVYKGRHKTTGQVVAMKKIRLESEEGVPSTAI-REISLLKELR-HP 61

QY 64 NIIYFNLDKIYDD-VILVTKLYRYDLSQLEIETKYCKRTTRFIYGINGNLVSNQYTLAN 122
Db 62 NIVS-LQVLMQDSRLYLIFEFLSMDLKKYLD-----SIPPGQY---- 100

QY 123 EIEEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIIFFARDDITQPIIGDFDICYDLKLP 182
Db 100 -MDSSLVKSILYQILQGIIVFCHSRVLRDLKPQNLLI--DDKGTIKLADFLGLARAFGI- 156

QY 183 PKDEPPMAKYID-VSTGIYKAPELILGITNYEYEDISLGIILTGLYSENFQSVLVKDD 241
Db 156 -----PIRVYTHYVTLWYRSPEVLLGSARYSTPVDIWSIGTIFAELATK----- 201

QY 242 KELTN-DSHVSDLYLLNQIFENFGTPNLTDFEDELFCDEYNENLHFKKFNQKYPRKDW 300
Db 201 KPLFHGDSIDQLF---RIFRALGTPNNEVWPEVESLQDYKNTFPKWKPGSLASHVKN-- 256

QY 301 DIILPRCNDLDMKEIFTKMIRYDRSKRITSKEIL 334
Db 256 -----LDENGLDLLSKMLIYDPAKRISGMAL 282
```

```
RESULT 13
US-08-461-985-16
; Sequence 16, Application US/08461985
; Patent No. 5872006
; GENERAL INFORMATION:
; APPLICANT: Boulton, Teri G.
; APPLICANT: Cobb, Melanie H.
; APPLICANT: Yancopoulos, George D.
; APPLICANT: Nye, Steven
; APPLICANT: Panayotatos, Nikos
; TITLE OF INVENTION: A Family of Map2 Protein Kinases
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```


STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,985
FILING DATE: 05-JUN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/176,620
FILING DATE: 03-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Mirock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 6526-123
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 353 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-461-985-14

Query Match 13.2%; Score 236.5; DB 2; Length 353;
Best Local Similarity 25.7%; Pred. No. 8.7e-14;
Matches 90; Conservative 62; Mismatches 125; Indels 73; Gaps 16;

QY	4	SDYYIDKELIYNSAISDIYTAIDKFNLPVCLKIVDEDFSLPPHSIH--REIFILKTLKP	61
Db	11	SDFQL-KSLLEGAYGVVCSATHKPTGEIVAICKI-EFFDKPLFALRTLREIKILKHEK-	68
QY	62	HPNIIIEYFN-----DLKIYDDVILVTKLYRYDLSQLIEITKYCKRTRTFIYGINGNLVSN	116
Db	68	HENIITIFNIQRPDSFENFNEVYIIQELMQTDLHRVIS-----	106
QY	117	QYTLANEIEEKDIKLWLSMSSGLEFIHSQGIHRDIKPSNIFF-ARDDITQPIIGDFDI	175
Db	106	-----TQMLSDDDHIQYFIYQTLRAVKVLHGSNVIHRDLKPSNLLINSNCDLK---VCDFGL	158
QY	176	-----CYDLKLPPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEDISLGIILTGL	228
Db	159	ARIIDESAADNSEPTGQSGMTEY--VATRWYRAPEVMLTSAKYSRAMDVWSCGILAE	216
QY	229	YSENFQSVLVKDDKELTNDSHVSDLYLLNQIFENFGTNPNTDFEDELFCDEYNNENLHFK	288
Db	217	FLR--RPFPGRDYR-----HQLLIFGIIGTPH---SDNDRCIESPAREYIK	261
QY	289	KENLQKYPRKWDIILPRCND---DLMKEIFTKMYDRSKRITSKEILQ	335
Db	262	--SLPMYPAAPLEKMFPRVNPKGIDLLQ----RMLVFDPAKRITAKEALE	305

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 1999, 06:24:57 ; Search time 41.67 Seconds
(without alignments)
6118.208 Million cell updates/sec

Title: US-09-072-994-13
Perfect score: 1019
Sequence: 1 ATGAAGTTGTCAGATTATTA.....TTACAATTAATGTTGGATTG 1019

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1007.8	98.9	1020	1	X27495
2	71.2	7.0	1671	1	Q24134
C 3	66.2	6.5	4590	1	N60472
C 4	64.2	6.3	1671	1	Q24134
C 5	60	5.9	110000	1	V21209_03
C 6	59.6	5.8	1864	1	N71405
C 7	58.6	5.8	605	1	T31530
C 8	58.2	5.7	19124	1	T72882
9	57.2	5.6	605	1	T31530
C 10	56.8	5.6	5760	1	N50530
C 11	56.6	5.6	5563	1	X08941
12	56.2	5.5	8920	1	Q62924
13	53.4	5.2	2503	1	Q53480
14	53.4	5.2	19124	1	T72882
15	53.2	5.2	9789	1	T41852
16	53	5.2	1070	1	T64450
17	53	5.2	2483	1	T63319
C 18	52.6	5.2	110000	1	V21209_15
C 19	51.8	5.1	51952	1	V26084
20	51.4	5.0	5203	1	V62901
C 21	51.4	5.0	110000	1	X20248_04
22	51.4	5.0	116277	1	X20249
C 23	51.2	5.0	110000	1	X20248_05
24	51	5.0	110000	1	V21209_13
25	51	5.0	110000	1	V21209_14
C 26	50.8	5.0	1230	1	N90223
27	50.8	5.0	4590	1	N60472
28	50.4	4.9	6124	1	Q03568
C 29	50.4	4.9	8920	1	Q62924
30	50.4	4.9	3974	1	T43360
C 31	50.2	4.9	10715	1	X20248_09
32	50	4.9	1644	1	V26286
33	50	4.9	3509	1	V26289
34	50	4.9	2574	1	V26290
35	50	4.9	2574	1	V26292
36	50	4.9	1711	1	V33136
37	50	4.9	3876	1	V30579
C 38	49.8	4.9	5852	1	Q11710
39	49.8	4.9	110000	1	X20248_06
40	49.8	4.9	111309	1	X20250
C 41	49.6	4.9	6124	1	Q03568
C 42	49.6	4.9	1711	1	V33136
C 43	49.4	4.8	1864	1	Q78892

44 49.2 4.8 2260 1 T27730 Insulin-stimulated
45 49.2 4.8 2260 1 T27731 Mutant insulin-sti

ALIGNMENTS

RESULT 1

X27495

ID X27495 standard; DNA; 1020 BP.

AC X27495;

DT 22-JUN-1999 (first entry)

DE C.albicans Cdk activating kinase (CAK) in vivo (CIV) 1 gene.

KW Protein kinase; cyclin-dependent kinase activating kinase; Cdk; CAK;

KW screening assay; fungicide; ss.

OS Candida albicans.

PN WO9907836-A1.

PD 18-FEB-1999.

PF 11-AUG-1998; F01788.

PR 12-AUG-1997; FR-010287.

PA (COMS) COMMISSARIAT ENERGIE ATOMIQUE.

PA (CURI-) INST CURIE.

PA (CNRS) CNRS CENT NAT RECH SCI.

PI Faye G, Mann C, Thuret JY, Valay JG;

DR WPI; 99-180489/15.

DR P-PSDB; W99578.

PT Candida albicans protein with cyclin-dependent kinase-activating

PT kinase activity - used to screen compounds for fungicidal and

PT antimycotic activity

PS Disclosure; Page 11-12; 19; French.

CC This sequence represents the coding region for a protein kinase of the

CC CIV1 family that has cyclin-dependent kinase (Cdk)-activating kinase

CC (CAK) activity that is not cyclin dependent. The new member designated

CC CAK in vivo (CIV1) differs from the ScCIV1 of Saccharomyces cerevisiae,

CC and lacks the motif: GxG(y/F)GxV, x = any amino acid. The sequence

CC can be used screening assays for fungicides, e.g. against Candida

CC albicans, or as industrial fungicides.

SQ Sequence 1020 BP; 385 A; 120 C; 149 G; 366 T;

Query Match 98.9%; Score 1007.8; DB 1; Length 1020;
Best Local Similarity 99.3%; Pred. NO. 6.2e-155;
Matches 1012; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ATGAAGTTGTCAGATTATTAATAGACAAGGAATTAATTTACAATAGTGCCATTCTGAT 60

Db 1 ATGAAGTTGTCAGATTATTAATAGACAAGGAATTAATTTACAATAGTGCCATTCTGAT 60

QY 61 ATATATACGGCTATTGATAAGTTTAATAACTTACCAGTATGCTTTAAATAGATTGATGAA 120

Db 61 ATATATACGGCTATTGATAAGTTTAATAACTTACCAGTATGCTTTAAATAGATTGATGAA 120

QY 121 GATTTTCAGTCTTCCACACATTCAATCCATCCGAGAAAATTTTATACCTAAAACCTTTGAAA 180

Db 121 GATTTTCAGTCTTCCACACATTCAATCCATCCGAGAAAATTTTATACCTAAAACCTTTGAAA 180

QY 181 CCACATCCAAACATAATGAATTTTAAATGATCTTAAATTTTATGATGATGATGATGATGATGAT 240

Db 181 CCACATCCAAACATAATGAATTTTAAATGATCTTAAATTTTATGATGATGATGATGATGATGAT 240

QY 241 GTCACCAAATTTGATCGTTATGATTTGAGTCAATTTGATTGAAATTTACAAAATATTGTAAA 300

Db 241 GTCACCAAATTTGATCGTTATGATTTGAGTCAATTTGATTGAAATTTACAAAATATTGTAAA 300

QY 301 CGAACACACACGATTTATTTATGGTATTAAATGGTAACTCTTGGTTAGTAATCAATATACACTT 360

Db 301 CGAACACACACGATTTATTTATGGTATTAAATGGTAACTCTTGGTTAGTAATCAATATACACTT 360

QY 361 GCTAATGAATTTGAAGAAAAGATATCAAAATATGGTTAAATCAATGAGTTTCAGGACTT 420

Db 361 GCTAATGAATTTGAAGAAAAGATATCAAAATATGGTTAAATCAATGAGTTTCAGGACTT 420

QY 421 GAATTTTATTCATTCACAAGGGGATAATTCATCGTGATATAAAACCCAGTAATATTTTCTTT 480

bad data
3
200 piny

Db 421 GAATTTATTCATCAAGGGATAATTCATCGTGATATAAAACCCAGTAATATTTCTTT 480
Qy 481 GCCCGGATGATATAACACACCGGATTTGAGATTTTGCATATTTGTTATGATTTAAAA 540
Db 481 GCCCGGATGATATAACACACCGGATTTGAGATTTTGCATATTTGTTATGATTTAAAA 540
Qy 541 CTGCCACCTAAAGATGAACCCCTATGCGGAAATATATTTGATGTATCTACAGGTATTTAT 600
Db 541 CTGCCACCTAAAGATGAACCCCTATGCGGAAATATATTTGATGTATCTACAGGTATTTAT 600
Qy 601 AAAGCACCAGAAATGATTTCTTGGTATAAATAATATGAATATGAATTTGGTCA 660
Db 601 AAAGCACCAGAAATGATTTCTTGGTATAAATAATATGAATATGAATTTGGTCA 660
Qy 661 TTGGGTATAATTTTGACTGGTTTATATTCAGAAAAATTTTCAAGTGTTTGTAGTCAAGAT 720
Db 661 TTGGGTATAATTTTGACTGGTTTATATTCAGAAAAATTTTCAAGTGTTTGTAGTCAAGAT 720
Qy 721 GATAAAGAAATGACTAATGATTTCTCATGTAGTGATTTATTTAATCAAAATATTT 780
Db 721 GATAAAGAAATGACTAATGATTTCTCATGTAGTGATTTATTTAATCAAAATATTT 780
Qy 781 GAAAAATTTCCGTACACCCCAATTTAACTGATTTTGAAGATGAATTTTGTGATGAATAT 840
Db 781 GAAAAATTTCCGTACACCCCAATTTAACTGATTTTGAAGATGAATTTTGTGATGAATAT 840
Qy 841 AATAATGAAAACTTGCATTTTAAAAAATTCAAATTTACAAAAATATCTAGAAAAAGATTGG 900
Db 841 AATAATGAAAACTTGCATTTTAAAAAATTCAAATTTACAAAAATATCTAGAAAAAGATTGG 900
Qy 901 GATATTATTTTACCTCGATGCAATGATGATTTAATGAAGAAAAATTTTACCAGATGATT 960
Db 901 GATATTATTTTACCTCGATGCAATGATGATTTTATGAAGAAAAATTTTACCAGATGATT 960
Qy 961 AGATATGATCGAAGTAAAGAAATAAATCTTAAAGAAATCTTAAAGAAATCTTAAAGTATG 1019
Db 961 AGATATGATCGAAGTAAAGAAATAAATCTTAAAGAAATCTTAAAGAAATCTTAAAGTATG 1019

RESULT 2

Q24134
ID Q24134 standard; DNA; 1671 BP.

AC Q24134;

DT 09-NOV-1992 (first entry)

DE 50 kD subunit of Scl.

KW Endonuclease Scl; PAGE; chromatography; ENS2 gene; ss.

OS Sacchromyces cerevisiae.

FH Key Location/Qualifiers

FT cds 1. .1589

FT /*tag= a

FT /label= SclI_50kd_subunit

FT exon 1. .81

FT /*tag= b

FT /number= 1

FT intron 82. .158

FT /*tag= c

FT /number= 1

FT exon 159. .1586

FT /*tag= d

FT /number= 2

PN J04104793-A.

PD 07-APR-1992.

PF 21-AUG-1990; 219566.

PR 21-AUG-1990; JP-219566.

PA (RIKA) RIKAGAKU KENKYUSHO.

DR WPI; 92-164281/20.

DR P-PSDB; R22667.

PT ENS2 gene encoding 50kD subunit of SclI endonuclease - used for mass

PT prodn. of recombinant endonuclease in eg. Sacchromyces IAM4274

PS Claim 1; Page 2; 4pp; Japanese.

CC The sequence given encodes the 50kD subunit of endonuclease SclI.

CC Endonuclease SclI can be mass produced from the 50kD subunits

CC expressed from the ENS2 gene. The 50kD subunit was purified by
CC treating SclI from eg. Sacchromyces IAM4274, and by denaturing the
CC protein and subjecting it to PAGE, or by chromatography on a
CC phosphocellulose column.
SQ Sequence 1671 BP; 769 A; 108 C; 113 G; 681 T;

Query Match (7.0%); Score 71.2; DB 1; Length 1671;

Best Local Similarity (45.6%); Pred. No. 0.00041;

Matches 365; Conservatve 0; Mismatches 428; Indels 7; Gaps 3;

QY 157 ATTTTATACCTTAAACCTTTGAAACCCACATCCAAACATAATTTGAATATTTTAATGATCTT 216

Db 23 AATCTTATGTTGACTTATCTTAACAGCATCATACTTAAAGATACATTATACCTACATT 82

QY 217 AAAATTTATGAT-GATGTTATATATTAGTCACCAAAATTTGTATCGTTATGATTTTGAGTCAATT 275

Db 83 AAATATATAAAATAAAATATATAAAAAAATAATAATAATAATAATAATAATAAAG 142

QY 276 GATTGAAATTAACAAATATTTGTAACGAACACACAGATTTTATTTATGTTATTAATGGTAA 335

Db 143 AATAAAAAATAAAACAATGAAAAACAAAAATTTAAATTTCTATTTTATTAATGTATATTA 202

QY 336 TCTTGTAGTAATCAATATACACTTGTCTAATGAAATTTGAAGAAAAAGATATCAAAATATG 395

Db 203 TTATATATTAATTTTAAATAATATTTTCATAAAAAATCAATTAATAAAAGACTGAATTAT 262

QY 396 GTTAAATCAATGAGTTGAGGACTTGAATTTATTTTATTCACAAAGGATATTTTCATCGTGA 455

Db 263 AGAATATGAATATATATAATAAATTTTAAATAAATAATAATAATAATAATAATAATA 322

QY 456 TATAAAACCCAGTAATTTTCTTGTCCCGGATGATATAAACAACCGATTTATTGGAGA 515

Db 323 TAATAATAAAATTTTATTTATTTATAGATATATATTAATGTTATATAAGTATCATATA 382

QY 516 TTTTGATATTTGTTATGATTTTAAACTGCCACCTAAAGATGAACCCCTATGGCGAATA 575

Db 383 ACAACGTACACCTATATCTAATAAAGATTAATAAATTTCAAAAAATATTATAG---ATTA 439

QY 576 TATTGATGATCTACAGGTATTTTATAAAGCACCAGAAATTTGTTGTTATTAACFAATTA 635

Db 440 TAAATTTATATATCTTTTATTTTATATTTTAAATAAATAAATAAATAAATAAATAA 499

QY 636 TGAATATGAAATTTGATATTTGGTCATTTGGSPATAAATTTTGACTGGTTTATATTCAGAAAA 695

Db 500 TAATAATAATAATAATAATAATTTTCATTTAAATAATAATAATGAATTAATAAATAATA 559

QY 696 TTTTCAAAAGTGTTTTAGTCAAAAGATGATATAA---GAATTTGACPAATGATTTCTCATGTAG 752

Db 560 TAATTTAAATTTATAAACTATCTAATATTTCAACTTAATTTTATCTAATAATTTTATTAAT 619

QY 753 TGATTTATATTTTAAATCAAAATATTTTGAAAAATTTTCGGTACACCCCAATTTTAACCTGATT 812

Db 620 AGATAAATATTTTAAATTAATAATAATAATAAATAATTTTAGATATATTAATAATAATTCCTAA 679

QY 813 TGAAGATGAATTTTGTGATGAATATAATAATAAATGAAAACTTGCAATTTTAAAAAATTCAA 872

Db 680 TAATATATATTTTAAATAATAATAATAATAAAGCTAAATTTAAATATTAATAAAGTATTAGA 739

QY 873 TTTACAAAAATATCCTAGAAAAAGATTTGGGATATTTTATTTTACCPCGATGCAATGATGATT 932

Db 740 TTTAAATAATAATGAATTTTATGATTTTATCAGGGTTAATTTGAAGGTGATGTTATAT 799

QY 933 AATGAAAGAAATTTTACCACCA 952

Db 800 TGGTCTCGGAGGTATTACAA 819

RESULT 3

N60472/c

ID N60472 standard; DNA; 4590 BP.

AC N60472;

DT 24-AUG-1991 (first entry)

Db	625	TTATCTATTAAATAAAATTTATTAGATAAAATTAAGTTGAATATTAGATAGTTTATAAATTT	566
QY	672	TTTGACTGGTTTATATTCAGAAAAATTTTCAAAGTGTTTTAGTCAAAAGATGATAAAGAATT	731
Db	565	AAATTATTATATATTTTAAATAATTCATATATTTTAAATGAATATTTATTTATTTA	506
QY	732	GACTAATGATTCATGTTAGTGATTTATATTTATTAATCAAAATATTTGAAAATTTTCGG	791
Db	505	TTATTATAAATTATCTATTCTATTTTATTTTATTTTAAATATATAAAATAAAGATATAAT	446
QY	792	TACACCCAATTTAACTGATTTTGAAGATGAATTTATTTTGTGATGA	836
Db	445	AATTATAATCTATAATATTTTTTTGAATTTATTAATCTTTTATTTA	401
RESULT 5			
V21209_03/c			
Continuation (4 of 17) of V21209 from base 300001 (Methanococcus jannaschii circular chr			
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209			
WP	Fragment Name	Begin	End
WP	V21209_00	1	110000
WP	V21209_01	100001	210000
WP	V21209_02	200001	310000
WP	V21209_03	300001	410000
WP	V21209_04	400001	510000
WP	V21209_05	500001	610000
WP	V21209_06	600001	710000
WP	V21209_07	700001	810000
WP	V21209_08	800001	910000
WP	V21209_09	900001	1010000
WP	V21209_10	1000001	1110000
WP	V21209_11	1100001	1210000
WP	V21209_12	1200001	1310000
WP	V21209_13	1300001	1410000
WP	V21209_14	1400001	1510000
WP	V21209_15	1500001	1610000
WP	V21209_16	1600001	1664976
Query Match 5.9%; Score 60; DB 1; Length 110000;			
Best Local Similarity 44.9%; Pred. No. 0.023;			
Matches 317; Conservative 0; Mismatches 380; Indels 9; Gaps 2;			
QY	261	TGATTGAGTCAATTGATTGAAATTACAAAATATTGTAACGGAACAACACGATTTATTTA	320
Db	85823	TGATATTTTGAATATTTATGAATGATTAAAGGATAAAGAACTAAAGGAGGTATATTGA	85764
QY	321	TGGTAFTAAATGGTAATCTTGTAGTAATCAATATACACTTGCTAATGAATGAATGAAGAAA	380
Db	85763	AAAAAAGATGATGAAATTTGTGTAATATAAATATAAAGTTGATATTTTATAATGTTGA	85704
QY	381	AGATATCAAAATATATGGTTAAATCAATAGTTCAGGACTTGAATTTATTCATTCACAAGG	440
Db	85703	ATTTAATGGTGATTTTAGATTTTGTAAATATGGAATATCAAAATTTGTGATATCCGTTTT	85644
QY	441	GATAATTCATCGTGATATAAAACCCAGTAATATTTTCTTGTCCCGGATGATATAACACA	500
Db	85643	TAATGGAAATGCTTATTTTAGAGTCATAACTTTTAAAGGAAGTGTTTATTTTATAAGAAC	85584
QY	501	ACCGATTATTGGAGATTTTGGATTTGTTA-----TGATTTAAAACTGCCACCTAAAGA	554
Db	85583	AATTTTAAATGGAGATGTTGATTTTATAGACACAAATTTTGAAGAAATGCTTATTTTAG	85524
QY	555	TGAACCCCTATAGCGAAATATATATGATGATCTACAGGTATTTTATAAGCACCAAGATT	614
Db	85523	TGTCACCGCTTTTAAAGGGAATATATTAAATTTTAGTGGCAACAATTTTAAATAAGAATC	85464
QY	615	GATTCCTGGTATAACTAATTATGAATATGAAATTTGATATTTGGTCATTTGGGTATAATTT	674
Db	85463	TCATTTTAAAGTACAACCTTTTGAAGGAAATACTTATTTTAGTGTCACAACCTTTTAATAT	85404
QY	675	GACTGGTTTATATTCAGAAAATTTTCAAAGTGTTTTAGTCAAAGATGATAAAGAATTGAC	734

Db	85403	AGCAGAATTTTATAAATTCACATTTAAATCTCATGTATATTTTGATGATATATCATTTAA	85344
QY	735	TAATGATTCCTCATGTTAGTGATTTATATATTTTAAATCAAAATATTTGAAAAATTTCCGTAC	794
Db	85343	TTTGTGTCTTTTACTGATTGTAGATTTAGAGATGATGTATCATTTAAAAAAATAGATAA	85284
QY	795	ACCCAATTTAACTGATTTTGAAGATGAATTTTGTGATGAATATAATAATGAAAACCT	854
Db	85283	AGAAAATTTTAAAGGTTTAGCAAT---ATTCTTAAAACTCAATTTCTAAACAACACAC	85227
QY	855	GCATTTTAAAAAATTCATTTTACAAAAATATCTAGAAAAAGATTGGGATATATTATTTACC	914
Db	85226	AACGATAGAAAACTTCCAATTATCAAAAAACATCATTTTTTAAAAACAGATGTTAGAGAAGT	85167
QY	915	TCGATGCAATGATGATTTAATGAAGAAATTTTACCAAGATGATT	960
Db	85166	GCTATTATGTGATGTTAAAAAAGAAGAAATTTTAAGTCATAAAAT	85121

RESULT 6

N71405/c

ID N71405 standard; DNA; 1864 BP.

AC N71405;

DT 18-APR-1991 (first entry)

DE Sequence of ANS-1 which increases transformation efficiency.

KW Enzyme; fungal expression vector; Aspergillus expression vector;

KW Trichoderma; ds.

OS Mucor miehei.

PN EP-215594-A.

PD 25-MAR-1987.

PF 27-AUG-1986; 306624.

PR 29-AUG-1985; US-771374.

PR 07-JUL-1986; US-882224.

PA (GENE-) GENENCOR INC.

PI Cullen D, Gray GL, Hayenga KJ, Lawlis VB;

DR WPI; 87-095049/14.

PT New DNA sequences for expressing polypeptide in filamentous fungi

PT - with secretion of prod. from the cells, and new vectors and

PT transformed fungi

PS Example; Fig 13; 45pp; English.

CC A DNA sequence coding for a heterologous polypeptide which can be

CC expressed in and secreted from filamentous fungi is claimed. Pref.

CC the DNA sequence codes for bovine preprochymosin, M. meihei

CC preprocarboxyl protease or A. niger preproglucoamylase. Also new

CC are vectors consisting of the DNA sequence plus an operably-linked

CC signal sequence. The vectors may also include a sequence which

CC increases transformation efficiency, e.g. ANS-1.

CC Sequence 1864 BP; 786 A; 210 C; 44 G; 732 T;

SQ

Query Match 5.8%; Score 59.6; DB 1; Length 1864;
Best Local Similarity 42.4%; Pred. No. 0.03;
Matches 397; Conservative 0; Mismatches 532; Indels 8; Gaps 3;

QY	86	ATAACTIACCAGTATGCTTAAAAATAGTTGATGAAGATTTTCAGTCTTCCACCACATTCAA	145
Db	1143	ATATCTGTNTAATATATTTANTATTTTATAAAATATCTAATTTATAGTAAAGTATAGTAGG	1084
QY	146	TCCATCGAGAAATTTTATATCTTAAACTTTTGAACACACATCCAAACATAATGAATATT	205
Db	1083	TTTTTATAATTTTATAATAATAATAAAATTTTAAAAAGTTATAAGNTAATTTNTAAANT	1024
QY	206	TTAATGATCTTAAATTTTATGATGATGTTATATATTAGTCACCAAAATGTATCGTTATGATT	265
Db	1023	TTTAACTTATTAATAATTAATAAATTTTATTATAAATTTTATAATTTATTTTGCAG	964
QY	266	TGAGTCAATTTGATTGAAATTTACAAAATATTGTAACGACACACACGATTTTATTTATGGTA	325
Db	963	GGAGAGTGTAAATAATAATAATAAATATAGTAGTTTATTATTATATATATAATA	904
QY	326	TTAATGCTAATCTTGTAGTAATCAATATACACTTCGTAATGAATTTGAAGAAAAAGATA	385
Db	903	ATATTTTNTAGTAGTGTAGTTTAATATAAGAGCTTTAAAAAATTTATTTCTTAATAGGAT	844

PT New malaria vaccines - contains cysteine-rich DBL family protein
PT binding domains homologous domains of the Duffy and sialic acid
PT binding proteins
PS Claim 4; Page 56-61; 96pp; English.
CC This sequence represents the var-7 gene of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see T72889 and T72888 respectively). The var
CC family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and SABP
CC mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the invention.
CC The compositions are for the treatment and prevention of malaria, and
CC comprise either a nucleotide sequence or encoded polypeptide of the
CC var-1, var-2, var-3 or var-7 genes of the DBL gene family, a family of
CC genes having homology with conserved regions of DABP and SABP. The
CC compositions are used for the treatment and prevention of malaria. They
CC are also used in the preparation of vaccines for inducing a protective
CC immune response in a mammal to Plasmodium merozoites (especially
CC Plasmodium falciparum or Plasmodium vivax).
CC Sequence 19124 BP; 7824 A; 2190 C; 2790 G; 6320 T;
SQ

Query Match 5.7%; Score 58.2; DB 1; Length 19124;
Best Local Similarity 45.2%; Pred. No. 0.047;
Matches 464; Conservative 0; Mismatches 543; Indels 20; Gaps 6;

QY 5 AGTTGTGAGATTAATATATAGACAAAGGAATTAATTTACAAATAGTGCCATTTCTGATATAT 64
DB 7195 AGTAGTATTATAATATGATGGTAGAATAAGAAATAAACACATTTTGTGAATGTATATATAT 7136
QY 65 ATACGGCTATTGATAAGTTTAAATAAACHTACCAGTATGCTTAAATAAGTTGATGAAGATT 124
DB 7135 GTAAGGTATAATTATGTATTACAATATATATAAATATTGTATATATATATATATATATA 7076
QY 125 TCAGTCTTCCACCACATTCATCCATCGAGAAATTTTATACCTAA-----AACCTTGAA 179
DB 7075 TTAATAGTTGACTATATAATATTACAATATATATGATATGCTTAAATAAATAATATTAA 7016
QY 180 ACCACATCCAAACATAATGAATATTTTAAATGATCTTAAATTTATGATGATGTTATATT 239
DB 7015 TATGATATAATAATAATATTATTAGTTTATATATTTTAAATAAATAATATATATATATT 6956
QY 240 AGTCACCAAAATTTGATCGTTATGATTTGAGTCAATTTGATGAAATACAAAATATTGTAA 299
DB 6955 AATAAAATTTATAATAAATTTAAATATTCTTACAAAATAAATAATATACAGAAATATTAT 6896
QY 300 ACGAACAACACGATTTATTTATGGTATTATTAATGGTAATCTTGTAGTAATCAATATACACT 359
DB 6895 ATTTTATGATTCCTTTATTTATCTATTTTAAATATATATATATATATATTTTATGTTTT 6836
QY 360 TGCTAATGAAATTTGAAGAAAAGATATCAAAATTTATGGTTAAATCAATGAGTTCAGGACT 419
DB 6835 ATTTATTAAGTAAATTTAATAATGAGAAAATAAATAACGAAAATACAAACATATAAATA 6776
QY 420 TGAATTTATTCATTCACAGGGATAATTCATCGTGATATAAACCAGTAATATTTTCTT 479
DB 6775 AGTATATATGCAACGTTTATATATATTTAATTAATTAACATTAATATATATATTTT 6716
QY 480 TGCCCGGATGATATAACACACCGGATTTATTTGGAGATTTTGTATATTTGTTATGATTTAA 539
DB 6715 TGACTTTATTTAATTTTATATATATATATATATATATATATAGAGATAACAAAAGAGACAAATA 6656
QY 540 ACTGCCACCTAAAGATGAACCCCTATGGCGAATATATTTGATGTATCTACAGGTATTTA 599
DB 6655 TGTTCTTCGTTCTCTCTATCTA-TATTATCTATCATATTTTATATATATATATATATA 6597
QY 600 TAAAGCACCAGAAATTTGATTCCTGGTATAACTAATTTATGAATATGAATTTGATTTGGTC 659
DB 6596 ATTGATATAGATACATATTTCTTGTATTTGTTATATATTTAAAGTAGTATATT----- 6541

QY 660 ATTGGGTATAAATTTTGACTGGTTTATATTTTCAAGAAATTTTCAAGTGTTTAGTCAAAGA 719
DB 6541 ATTATTATAAATTTTGTGTTATATATTTATTAATTTATATATAATAATATATATAGCAT 6482
QY 720 TGATAAAGAAATGACTAATGATCTCTCATGTTAGTCATTTATTTTAAATCAATATT 779
DB 6481 CAAAAAATAATGATAAATAAATAACAGGAAATAATATATATATATATATATATATAT 6422
QY 780 -----TGAAATTTCCGGTACACCCAAATTTAACTGATTTTGAAGATGAATTTT 831
DB 6421 ATTAATAAATAATGTTTATCATTTGTTTGTGTTATTTTATGTTATTCATGCATTT 6362
QY 832 GATGAATATAATAAATGAAAACTTGCATTTTAAATAATCAATTTTACAAAATATCTAGA 891
DB 6361 TATGAATTTCAAAATTTTATTGTATAATAATAAATAAATAAAGTAAATAACACATATA 6302
QY 892 AAAGATTGGGATATTATTTTACCTCGATGCAATGATGATTAATGAAGAAATTTTACC 951
DB 6301 TATATATATTCAAAATATGAGTTATTAATAAATAAATGTTCAATGTTCTATATATATATA 6242
QY 952 AAGATGATTAGATATGATCGAAGTAAAGAAATAAATCTTAAAGAAATCTTACAATTAATG 1011
DB 6241 GAAATATTTGTTATATAATAATAAATACATATAT-GCTACTATATAAATAATTAATAATATCT 6183
QY 1012 TTGGATT 1018
DB 6182 TTAAAGT 6176

RESULT 9
T31530
ID T31530 standard; cDNA; 605 BP.
AC T31530;
DT 15-SEP-1996 (first entry)
DE Human 3' apolipoprotein B SAR element clone Rh32.
KW Erythropoietin; EPO; anaemia; gene therapy; vector;
KW scaffold attachment region; SAR element; apolipoprotein B;
KW transgenic animal; ss.
CS Homo sapiens.
FN WO9619573-A1.
PD 27-JUN-1996.
PF 18-DEC-1995; CA0696.
FR 19-DEC-1994; US-358918.
PA (CANG-) CANGENE CORP.
PI Delcuve G;
DR WPI; 96-309587/31.
PT Recombinant DNA molecule expressing mammalian erythropoietin -
PT useful to transform cell lines, and for gene therapy, e.g. of
PT anaemia and other red blood cell disorders
PT Claim 7; Page 59-60; 84pp; English.
PS Human apolipoprotein B (apoB) scaffold attachment region (SAR)
CC element clones Rh32 (T31530) and Rh10 (T31531) respectively carry
CC the 3' human apoB SAR element and the distal 1212 bp 5' human apoB
CC SAR element and 1317 bp proximal sequence. These SAR elements
CC co-map with the boundaries of the human apoB gene chromatin
CC domain. A novel recombinant DNA molecule adapted for transfection
CC of a host cell comprises an erythropoietin (EPO) cDNA (T31529) or
CC genomic clone (T31532) operably linked to an expression control
CC sequence and to the 5' and 3' SAR elements. The SAR elements
CC increase expression of the recombinant EPO in stable, long-term
CC mammalian cell cultures.
SQ Sequence 605 BP; 278 A; 18 C; 35 G; 274 T;

Query Match 5.6%; Score 57.2; DB 1; Length 605;
Best Local Similarity 46.3%; Pred. No. 0.077;
Matches 188; Conservative 0; Mismatches 218; Indels 0; Gaps 0;

QY 571 AAATATATTGATGTATCTACAGGTATTTATAAAGCACCAGAAATGATTTGGTATAACT 630
DB 63 AAATATTTATAAATAATTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 122

• Sat Nov 6 11:07:18 1999

us-09-072-994-13.rng

Page 11

Search completed: November 6, 1999, 07:50:24
Job time: 5127 sec

QY 118 YTLANEIEEKDIKWLKSMSSGLEFIHSQGIHHRDIKPSNIFFARDITQPIIGDFDICY 177
Db 112 -----PLETQFIQFLYQILRGLKFKVHSAGVIHRDLKPSNILINEN-----CDLKIC- 159
QY 178 DLKLPKDEPPMAKYIDVSTGIYKAPELILGITNYEYEDISLGLITGLYSENFQSVL 237
Db 159 DFGLARIQDPMQMGY--VSTRYRAPEIMLTWQYNVEVDIWSAGCIF-----AEM 207
QY 238 VKDDKELTNDSHVSDLYLLNQIFENFGTGNLTDPEDELFCDEY-----NNENLHFK 288
Db 208 IEGKPLFPGRDHVNQFSIITEL--LGTPPMEVI--ETICKNTLRFVQSLPQKEKVPF- 262
QY 289 KFNLOKYPKRDWDIILPRCNDLDMKEIFTKMIYDRSKRITSKEIL 334
Db 262 ---AEKFNADPDAI-----DLLEKMLVDFPRKRISAADAL 294

RESULT 2
S24386
protein kinase (EC 2.7.1.37) cdc2 homolog - slime mold (Dictyostelium discoideum)
C;Species: Dictyostelium discoideum
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Sep-1997
C;Accession: S24386
R;Michaelis, C.; Weeks, G.
Biochim. Biophys. Acta 1132, 35-42, 1992
A;Title: Isolation and characterization of a cdc2 cDNA from Dictyostelium discoideum.
A;Reference number: S24386; MUID:92379089
A;Accession: S24386
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-296 <MIC>
A;Cross-references: EMBL:M80808; NID:g167685; PID:g167686
C;Superfamily: kinase-related transforming protein; protein kinase homology
C;Keywords: ATP; phosphoprotein; phosphotransferase; serine/threonine-specific protein kinase
F:8-257/Domain: protein kinase homology <KIN>
F:16-24/Region: protein kinase ATP-binding motif
F:39,55,129,131/Active site: Lys, Glu, Asp, Lys #status predicted

Query Match 15.2%; Score 273; DB 2; Length 296;
Best Local Similarity 26.0%; Pred. No. 2e-11;
Matches 86; Conservative 63; Mismatches 122; Indels 60; Gaps 13;

QY 7 YIDKELIYNSAIDYTAIDKFNLPVCLKIVD-EDFSLPPHSIHREIFILKTLKPHNI 65
Db 10 YQKLEKLGEGYGVYKAKKATGRMVALKIRLEDGVPSTAL-REISLLKEV-PHPNV 67
QY 66 IEYFNDLKIYDDVILVTKLYRYDLSOLIE-ITKYCKRTTRFIYINGNLVSNQYTLANEI 124
Db 68 VSLFDVLHCQNRLYLVFEYLDQDLKYMDSVPALCPQL----- 106
QY 125 EEKDIKWLKSMSSGLEFIHSQGIHHRDIKPSNIFFARDITQPIIGDFDICYDLKLPK 184
Db 106 ----IKSYLYQLLGLAYSHGHRILHRDLKPQNLLIDRQALK--LADFLARAVSIPVR 159
QY 185 DEPPMAKYIDVSTGIYKAPELILGITNYEYEDISLGLITGLYSENFQSVLVKDDREL 244
Db 160 -----VYTHEIVTLWYRAPEVLLGSKSYSPVDMWSVGC-----FGEMLNK--KPL 204
QY 245 -TNDSHVSDLYLLNQIFENFGTGNLTDPEDELFCDEYNNENLHFKFNLOKYPKRDWDII 303
Db 205 FSGDCEIDQIF--RIFRVLGTPDDSIWPGVKLPEYVS-----TFPNWPGQPYNKI 253
QY 304 LPRCNDLDMKEIFTKMIYDRSKRITSKEIL 334
Db 254 FPRC-EPLALDLIAKMLQYEPSKRISAKEAL 283

RESULT 3
S56225
hypothetical protein yFL029c - yeast (Saccharomyces cerevisiae)
C;Species: Saccharomyces cerevisiae
C;Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 06-Feb-1998

C;Accession: S56225
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sas
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces
A;Reference number: S56186
A;Accession: S56225
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-368 <MUR>
A;Cross-references: EMBL:D50617; NID:g836685; PID:d1009850; PID:g836725; MIPS:YFL029c
C;Genetics:
A;Gene: SGD:CAK1
A;Cross-references: SGD:S0001865; MIPS:YFL029c
A;Map position: 6L

Query Match 14.9%; Score 267; DB 2; Length 368;
Best Local Similarity 25.1%; Pred. No. 6.4e-11;
Matches 98; Conservative 66; Mismatches 144; Indels 82; Gaps 14;

QY 1 MKLSDYYIDK-ELIYNSAIDYTAIDKFNLPVCLKIVDEDFSLPPHSIHREIFILKTL 59
Db 1 MKLSDIDITHCQLVKSTRTARIYRS-DTY-----AIKCLALDFDIPPHNAKFEVSILNKL 54
QY 60 KPHNIIIEYFNDLKIYD--DVILVTKLYRYDLSQLEITKYCKRTTRF----- 106
Db 55 GNRCKHILPLLESKATDNDLLLLFPFEEMNLVEFMQM--HYKRRDRKKNPYDYDLLNPSI 112
QY 106 -----IYGINGLVSNQYTLANEIEEKDIKWLKSMSSGLEFIHSQGIHHRDIKPSNI 158
Db 113 PIVADPPVQKYTNQLDVNRYSL-----FFRQMVGEIAFLHENKIIHRDIKQNI 162
QY 159 FFARDIT---OPIIGDFDICYDL--KLPPKDEPPMAKYIDVSTGIYKAPELILGITNYE 213
Db 163 MLTNTSTVSPKLYIIDFGISYDMANNSSQTSAPEDMSKVTDISTGIYKAPELVFGVKCYD 222
QY 214 YEIDIWSLGIILTGLYSENFQSV---LVKDDKELTNDSHVSDLYLLNQIFENFGTGNLT 269
Db 223 GGVDVWSLLIIISQWFQRETSRMGHVPAMIDGSDMNSDGSDFRLICSIKELGIPSIQ 282
QY 270 DFED-----ELFCDEYNNENLHFKFNLOKYPKRDWDIILPRCND 309
Db 283 KWEEVAQHGSVDFAVGMFGADGDKYVLDQEKDVQISIVERNM---PRLD-----EIAD 333
QY 310 DLMKEIFTK---MIRYDRSKRITSKEILQ 335
Db 334 VKVKQKFINCILGMVSFSPNERWSCQRILO 363

RESULT 4
S53538
protein kinase (EC 2.7.1.37) cdc2 homolog - Paramecium tetraurelia (SGC5)
C;Species: Paramecium tetraurelia
C;Date: 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 28-Feb-1997
C;Accession: S53538
R;Tang, L.; Pelech, S.L.; Berger, J.D.
Biochim. Biophys. Acta 1265, 161-167, 1995
A;Title: Isolation of the cell cycle control gene cdc2 from Paramecium tetraurelia.
A;Reference number: S53538; MUID:95210349
A;Accession: S53538
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-308 <TAN>
A;Note: the sequence from Fig. 3 is inconsistent with that from Fig. 1 in having 70-A
C;Genetics:
A;Genetic code: SGC5
C;Superfamily: kinase-related transforming protein; protein kinase homology
C;Keywords: ATP; phosphotransferase
F:7-262/Domain: protein kinase homology <KIN>
F:15-23/Region: protein kinase ATP-binding motif
F:38,56,133,135/Active site: Lys, Glu, Asp, Lys #status predicted

QY 65 IIEYFNDLKIYDDVILVTKLRYDLSQLIEITKYCKRTTRFIYINGNLVSNQYTLANEI 124
| | : : | | | : : : | | : : | | : :
Db 94 IITLADIVK--DPV-----SRTPALVFEHVNTDFKQ--LYQTL 128
| | : : | | : : | | : : | | : : | | : :
QY 125 EEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYDLKLPK 184
: | | : : : | | : : | | : : | | : : | | : :
Db 129 TDYDIRFYMEILKALDYCHSMGIMHRDVKPHNVMDHEHRKLRLI-----DWGLAEF 181
| | : : | | : : | | : : | | : : | | : :
QY 185 DEPPMAKYIDVSTGIYKAPELILGITNYEYEIDISLGIILTGLY--SENFQSVLVKDDK 242
| | : : | | : : | | : : | | : : | | : :
Db 182 YHPGQEQYNNRVASRYFKGPELLVDYQMYDYSLDWMSLGCMLASMIFFRKEPF----- 233
| | : : | | : : | | : : | | : : | | : :
QY 243 ELTNDSHVSDLY-LLNQIFENFGTNPNTDFEDELFCDEYNNENLHFKKFN--LQKYPRKD 299
| | : : | | : : | | : : | | : : | | : :
Db 233 -----FHGHDNYDQLVRIAKVLGTEDLYD-----YIDKYNIE--LDPFRNDILGRHSRKR 280
| | : : | | : : | | : : | | : : | | : :
QY 300 WDILPRCNDLDMK---EIFTKMIRYDRSKRITSKEILQ 335
| | : : | | : : | | : : | | : : | | : :
Db 281 WERFVHSENQHLVSPALDFLDKLLRYDHQSRLTAREAME 320
| | : : | | : : | | : : | | : : | | : :
RESULT 7
S21335
casein kinase II (EC 2.7.1.1-) alpha-1 chain - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 19-Dec-1997
C:Accession: S21335; A35206
R;Watanabe, M.; Yuge, M.; Maeda, O.; Ohno, S.; Kawasaki, H.; Suzuki, K.; Hidaka, H.
submitted to the EMBL Data Library, October 1990
A:Description: Nucleotide sequence of cDNA for casein kinase II alpha subunit from bovine
A:Reference number: S21335
A:Accession: S21335
A:Molecule type: mRNA
A:Residues: 1-391 <WAT>
A:Cross-references: EMBL:X54962; NID:g610; PID:g611
A:Experimental source: testis
R;Litchfield, D.W.; Lozeman, F.J.; Piening, C.; Sommercorn, J.; Takio, K.; Walsh, K.A.;
J. Biol. Chem. 265, 7638-7644, 1990
A:Title: Subunit structure of casein kinase II from bovine testis. Demonstration that th
A:Reference number: A35206; MUID:90237072
A:Accession: A35206
A>Status: preliminary
A:Molecule type: protein
A:Residues: 50-59;103-122;230-239;248-279;284-303 <LIT>
C:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: ATP; heterotetramer; phosphotransferase; serine/threonine-specific protein k
F:37-299/Domain: protein kinase homology <KIN>
F:45-53/Region: protein kinase ATP-binding motif

Query Match 14.6%; Score 262; DB 2; Length 391;
Best Local Similarity 26.5%; Pred. NO. 1.5e-10;
Matches 90; Conservative 59; Mismatches 125; Indels 66; Gaps 13;

QY 5 DYIDKELIYNSAISDIYTAIDKFNNLPVCLKIYDEDFSLPPHSIHREIFILKTLKPHNP 64
| | : : | | : : | | : : | | : : | | : : | | : :
Db 38 DYQLVRKL-GRGKYSEVFEAINITNNEKVVVKILK---PVKKKIKIKIKILENLRGGPN 93
| | : : | | : : | | : : | | : : | | : : | | : :
QY 65 IIEYFNDLKIYDDVILVTKLRYDLSQLIEITKYCKRTTRFIYINGNLVSNQYTLANEI 124
| | : : | | : : | | : : | | : : | | : : | | : :
Db 94 IITLADIVK--DPV-----SRTPALVFEHVNTDFKQ--LYQTL 128
| | : : | | : : | | : : | | : : | | : : | | : :
QY 125 EEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYDLKLPK 184
: | | : : : | | : : | | : : | | : : | | : : | | : :
Db 129 TDYDIRFYMEILKALDYCHSMGIMHRDVKPHNVMDHEHRKLRLI-----DWGLAEF 181
| | : : | | : : | | : : | | : : | | : : | | : :
QY 185 DEPPMAKYIDVSTGIYKAPELILGITNYEYEIDISLGIILTGLY--SENFQSVLVKDDK 242
| | : : | | : : | | : : | | : : | | : : | | : :
Db 182 YHPGQEQYNNRVASRYFKGPELLVDYQMYDYSLDWMSLGCMLASMIFFRKEPF----- 233
| | : : | | : : | | : : | | : : | | : : | | : :
QY 243 ELTNDSHVSDLY-LLNQIFENFGTNPNTDFEDELFCDEYNNENLHFKKFN--LQKYPRKD 299
| | : : | | : : | | : : | | : : | | : : | | : :
Db 233 -----FHGHDNYDQLVRIAKVLGTEDLYD-----YIDKYNIE--LDPFRNDILGRHSRKR 280
| | : : | | : : | | : : | | : : | | : : | | : :

QY 300 WDILPRCNDLDMK---EIFTKMIRYDRSKRITSKEILQ 335
| | : : | | : : | | : : | | : : | | : : | | : :
Db 281 WERFVHSENQHLVSPALDFLDKLLRYDHQSRLTAREAME 320
| | : : | | : : | | : : | | : : | | : : | | : :
RESULT 8
B30319
casein kinase II (EC 2.7.1.1-) alpha chain - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 28-Feb-1997
C:Accession: B30319
R;Meisner, H.; Heller-Harrison, R.; Buxton, J.; Czech, M.P.
Biochemistry 28, 4072-4076, 1989
A:Title: Molecular cloning of the human casein kinase II alpha-subunit.
A:Reference number: A90543; MUID:89323123
A:Accession: B30319
A:Molecule type: mRNA
A:Residues: 1-384 <MEI>
C:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: ATP; heterotetramer; phosphotransferase; serine/threonine-specific protei
F:30-292/Domain: protein kinase homology <KIN>
F:38-45/Region: protein kinase ATP-binding motif

Query Match 14.6%; Score 262; DB 2; Length 384;
Best Local Similarity 26.5%; Pred. No. 1.4e-10;
Matches 90; Conservative 59; Mismatches 125; Indels 66; Gaps 13;

QY 5 DYIDKELIYNSAISDIYTAIDKFNNLPVCLKIYDEDFSLPPHSIHREIFILKTLKPHNP 64
| | : : | | : : | | : : | | : : | | : : | | : :
Db 31 DYQLVRKL-GRGKYSEVFEAINITNNEKVVVKILK---PVKKKIKIKIKILENLRGGPN 86
| | : : | | : : | | : : | | : : | | : : | | : :
QY 65 IIEYFNDLKIYDDVILVTKLRYDLSQLIEITKYCKRTTRFIYINGNLVSNQYTLANEI 124
| | : : | | : : | | : : | | : : | | : : | | : :
Db 87 IITLADIVK--DPV-----SRTPALVFEHVNTDFKQ--LYQTL 121
| | : : | | : : | | : : | | : : | | : : | | : :
QY 125 EEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDICYDLKLPK 184
: | | : : : | | : : | | : : | | : : | | : : | | : :
Db 122 TDYDIRFYMEILKALDYCHSMGIMHRDVKPHNVMDHEHRKLRLI-----DWGLAEF 174
| | : : | | : : | | : : | | : : | | : : | | : :
QY 185 DEPPMAKYIDVSTGIYKAPELILGITNYEYEIDISLGIILTGLY--SENFQSVLVKDDK 242
| | : : | | : : | | : : | | : : | | : : | | : :
Db 175 YHPGQEQYNNRVASRYFKGPELLVDYQMYDYSLDWMSLGCMLASMIFFRKEPF----- 226
| | : : | | : : | | : : | | : : | | : : | | : :
QY 243 ELTNDSHVSDLY-LLNQIFENFGTNPNTDFEDELFCDEYNNENLHFKKFN--LQKYPRKD 299
| | : : | | : : | | : : | | : : | | : : | | : :
Db 226 -----FHGHDNYDQLVRIAKVLGTEDLYD-----YIDKYNIE--LDPFRNDILGRHSRKR 273
| | : : | | : : | | : : | | : : | | : : | | : :

QY 300 WDILPRCNDLDMK---EIFTKMIRYDRSKRITSKEILQ 335
| | : : | | : : | | : : | | : : | | : : | | : :
Db 274 WERFVHSENQHLVSPALDFLDKLLRYDHQSRLTAREAME 313
| | : : | | : : | | : : | | : : | | : : | | : :

RESULT 9
A38611
casein kinase II (EC 2.7.1.1-) alpha chain - chicken
C:Species: Gallus gallus (chicken)
C:Date: 23-Aug-1991 #sequence_revision 23-Aug-1991 #text_change 31-Oct-1997
C:Accession: A38611
R;Maridor, G.; Park, W.; Krek, W.; Nigg, E.A.
J. Biol. Chem. 266, 2362-2368, 1991
A:Title: Casein kinase II. cDNA sequences, developmental expression, and tissue distr
A:Reference number: A38611; MUID:91115855
A:Accession: A38611
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-391 <MAR>
A:Cross-references: GB:M59456; GB:J05736; NID:g211533; PID:g211534
C:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: ATP; heterotetramer; phosphotransferase; serine/threonine-specific protei
F:37-299/Domain: protein kinase homology <KIN>
F:45-53/Region: protein kinase ATP-binding motif

Query Match	14.6%	Score 262;	DB 2;	Length 391;
Best Local Similarity	26.5%;	Pred. No. 1.5e-10;		
Matches 90;	Conservative 59;	Mismatches 125;	Indels 66;	Gaps 13;

QY	5	DYIDKELIYNSAISDIYTAIDKFNLPVCLKIVDEFSLPPHSIHREIFILKTLKPHPN	64
Db	38	DYQLVRKL-GRGKYSEVFEAINITNNEKVVVKILK---PVKKKKIKREIKILENLRGSPN	93
QY	65	IIIEYFNOLKIYDDVILVTKLYRYDLSQLIEITKYCKRTTRFYINGNLSVSNQYTLANEI	124
Db	94	IITLADIVK--DPV-----SRTPALVFEHVNTDFKQ--LYQTL	128
QY	125	EEDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDITQPIIGDFDICYDKLPPK	184
Db	129	TDYDIRFYMEILKALDYCHSMGIMHRDVKPHNMIDHEHKKRLRI-----DWGLAEF	181
QY	185	DEPPMAKYIDVSTGIYKAPELILGITNYEYEIDIWSLGIILTGLY--SENFQSVLVKDDK	242
Db	182	YHPGQEYNYRVASRYFKGPPELLVDYQMYDYSLDMWSLGCMLASMIFRKEPF-----	233
QY	243	ELTNDSHVSDLY-LLNQIFENFGTPNLTDFEDELFCDEVNNENLHFKFN--LQYPRKD	299
Db	233	-----FHCHDNYDQLVRIAKVLGTEDLYD-----YIDKYNIE--LDPRFNDILGRHSKR	280
QY	300	WDIILPCRNDDLMK----EIFTKMIRYDRSKRITSKEILQ	335
Db	281	WERFVHSENQHLVSPEALDFDLKLRDYDHQSRLTAREAME	320

RESULT 10
A41227
protein kinase (EC 2.7.1.37) cdk2 - human
N:Alternate names: Egl homolog; protein kinase p34
C:Species: Homo sapiens (man)
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 24-Sep-1998
C:Accession: A41227; S17873; S16520
R:Ninomiya-Tsujii, J.; Nomoto, S.; Yasuda, H.; Reed, S.I.; Matsumoto, K.
Proc. Natl. Acad. Sci. U.S.A. 88, 9006-9010, 1991
A:Title: Cloning of a human cDNA encoding a CDC2-related kinase by complementation of a
A:Reference number: A41227; MUID:92020980
A:Accession: A41227
A:Molecule type: mRNA
A:Residues: 1-298 <NIN>
A:Cross-references: GB:M68520; NID:g180177; PID:g180178
R:Tsai, L.H.; Harlow, E.; Meyerson, M.
Nature 353, 174-177, 1991
A:Title: Isolation of the human cdk2 gene that encodes the cyclin A- and adenovirus E1A
A:Reference number: S17873; MUID:91367262
A:Accession: S17873
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-298 <TSA>
R:Elledge, S.J.; Spottswood, M.R.
EMBO J. 10, 2653-2659, 1991
A:Title: A new human p34 protein kinase, CDK2, identified by complementation of a cdc28
A:Reference number: S16520; MUID:91330891
A:Accession: S16520
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-176, 'S', 178-298 <ELL>
A:Cross-references: EMBL:X61622; NID:g29848; PID:g29849
C:Genetics:
A:Gene: GDB:CDK2
A:Cross-references: GDB:128984; OMIM:116953
A:Map position: 12ql3-12ql3
C:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: ATP; cell cycle control; mitosis; phosphoprotein; phosphotransferase; serine
F:2-255/Domain: protein kinase homology <KIN>
F:10-18/Region: protein kinase homology
F:14,160/Binding site: phosphate (Thr) (covalent) #status predicted
F:15/Binding site: phosphate (Tyr) (covalent)#status predicted

```
F;33,51,127,129/Active site: Lys, Glu, Asp, Lys #status predicted
```

	Query Match	14.6%;	Score 262;	DB 2;	Length 298;
	Best Local Similarity	26.2%;	Pred. No. 1.1e-10;		
	Matches 83;	Conservative 59;	Mismatches 119;	Indels 56;	Gaps 11

QY	21	IYTAIDKFNNLPVCLKIV--DEDFSLPPHSIHREIFILKTLKPHPNIEFYNDLKIYDDV	78
	:	: : :	:
Db	18	VYKARNKLGTGEVALKKIRLDTEGVPSTAIRESLLKELN-HPNIVKLLDVIHTENKL	76
	:	: : :	:
QY	79	ILVTKLYRYDLSQLIEITKYCKRTTRFIYGINGNLVSQYTLANEIEEKDIKLWLKSMSS	138
	:	: : :	:
Db	77	YLVEEFLHQDLKKFMDASA-----LTGIPLPL-----IKSYLFQLLQ	113
	:	: : :	:
QY	139	GLEFIHSGIIRHDIKPSNIFFARDDITQPIIGDFDICVDLKPPKDEPPMAKYI-DVST	197
	:	: : :	:
Db	114	GLAFCHSHRVHLRDLKPQNLLINTEGAIK--LADEGLARAFGV-----PVRTYTHEVVT	165
	:	: : :	:
QY	198	GIYKAPELILGITNYEYEIDIWSLGIIITGLYSENFQSVLVKDDKELTNSHVS DLYLLN	257
	:	: : :	:
Db	166	LWYRAPEILLGCKYYSTAVIDWSLGCIF-----AEMVTRRALFPGDSEIDQLF---	214
	:	: : :	:
QY	258	QIFENFGTPNLTDFFEDELFCDEYNNEHLHKFKFNLOKYPKDKWDIILPRCNDDDLMEIFT	317
	:	: : :	:
Db	214	RIFRTLGTDPDEVVWPVGVTSPMPDY-----KPSFPKWARKQDFSKVVVPLDED-GRSLLS	264
	:	: : :	:
QY	318	KMIRYDRSKRITSKEIL	334
	:	: : :	:
Db	265	QMLHYDPNKRI SAKAAL	281

RESULT 11
A44878
protein kinase (EC 2.7.1.37) cdk2 - goldfish
C:Species: Carassius auratus (goldfish)
C:Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 20-Mar-1998
C:Accession: A44878
R:Hirai, T.; Yamashita, M.; Yoshikuni, M.; Tokumoto, T.; Kajiura, H.; Sakai, N.
Dev. Biol. 152, 113-120, 1992
A:Title: Isolation and characterization of goldfish cdk2, a cognate variant of
A:Reference number: A44878; MUID:92331802
A:Accession: A44878
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-298 <HIR>
A:Cross-references: GB:S40289; NID:g251619; PID:g251620
A:Experimental source: oocyte
A:Note: sequence extracted from NCBI backbone (NCBIN:108782, NCBIP:108783)
C:Genetics:
A:Gene: cdk2
C:Superfamily: kinase-related transforming protein; protein kinase homology
C:Keywords: ATP; cell cycle control; mitosis; phosphoprotein; phosphotransferase
F:2-255/Domain: protein kinase homology <KIN>
F:10-18/Region: protein kinase ATP-binding motif
F:33,51,127,129/Active site: Lys, Glu, Asp, Lys #status predicted

	Query Match	14.6%;	Score 261;	DB 2;	Length 298;
	Best Local Similarity	27.3%;	Pred. No. 1.2e-10;		
	Matches	87; Conservative	61; Mismatches	111; Indels	60; Gaps
					14;
QY	21	IYTAIDKFNPNLPVCLKIV--DEDFSLPPHSIHREIFILTKLPHPNIIIEYFNDLKIIDV	78		
		: : : : : : : : : :			
Dd	18	VYKAKNKVTGETVALKKIRLDTEGVPSTAIREISLLKELN-HPNIVKLHDVIHTENKL	76		
QY	79	ILVTKLYRYDLSQLIETIKYCKRTRTFIYGINGNLVSNOYTLANEIEEKDIKLWLKSMSS	138		
		: : : : : : : : : :			
Dd	77	YLVEFEFLHQDLKRFD-----SSTVTGISPLV-----KSYLEQLLQ	113		
QY	139	GLEFIHSGQIIHRDIKPSNIFF-ARDDITQPIIGDFDICYDLKLPPKDEPPMAKYI-DVS	196		
		: : : : : : : :			
Dd	114	GLAFCHSHRVLHRDLKPQNLLINAQEIK---LADFGLARAFVG-----PVRTYTHEW	164		

QY	60	KPHNIIIEYN----	DLKIYDDVILTKLYRYDLSQLLEITKYCKRTRFRFYINGNLVS	115
		: : : : : : :		
Db	78	H-HPNIVNLLDVICADGKLY----	LIFEYVDHDLKKALE-----KRGGA	118
QY	116	QNYTLANEIEEKDIKLWLKSMSSGLEFIHSQGIHRDIKPSNIFFARDDITQPIIGDFDI	175	
		: : : : : : : : : : :		
Db	118	-----TGTTLKKIIYQLLEGISFCHRRHIVRDLUKPANILVTTDNSVK--IADFG	166	
QY	176	CYDLKLPKDEPPMAKYI-DVSTGIYKAPELILGITNYEYEIDIWLSGIILTGLYSENFQ	234	
		: : : : : : : : : : : : : :		
Db	167	ARAFQI-----PMHTYTHEVTVIWMYRAPEILLGEKHYPAPVDMWSIGCIAELAR--G	217	
QY	235	SVLVKDDKELTNDSHVSDLYLLNQIFENFGTP-----NLTDFELELFCDEYNNE	283	
		: : : : : : : :		
Db	218	KVLFGRGDSEICQ-----LFEIIFOVLGTPMDAEGSWLGVSLLPDYRDV-----	260	
QY	284	NLHFKKFNLQYPRKDWDIILPRCNDDDLKMEIFTKMTIYDRSKRITSKEILQ	335	
		: : : : : : : : :		
Db	260	-----FPKWSGKPLTQVLPTLDGDAV-DLLSQMLRYNPAERISAKAALQ	302	

RESULT 15

c-Jun amino-terminal kinase (EC 2.7.1.-) 2 alpha1 - chicken
 N;Alternate names: stress-activated protein kinase
 C;Species: Gallus gallus (chicken)
 C;Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 10-Jul-1998
 C;Accession: JC5531
 R;Ishikawa, T.; Nakada-Moriya, Y.; Ando, C.; Tanda, N.; Nishida, S.; Minatogawa, Y.; Noh
 Biochem. Biophys. Res. Commun. 234, 489-492, 1997
 A;Title: Expression of the JNK2-alpha1 gene in the developing chick brain.
 A;Reference number: JC5531; MUID:97320448
 A;Accession: JC5531
 A;Status: nucleic acid sequence not shown
 A;Molecule type: DNA
 A;Residues: 1-382 <ISH>
 A;Cross-references: DDBJ:AB000807; NID:g1816447; PID:d1019937; PID:g1816448
 C;Comment: This enzyme is involved in development of the central nervous system as a med
 actors. It plays a role in the differentiation of central nervous system.
 C;Genetics:
 A;Gene: JNK2-alpha1
 C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog
 C;Keywords: phosphotransferase
 F;24-278/Domain: protein kinase homology <KIN>
 F;230/Active site: His #status predicted

Query Match 14.5%; Score 259.5; DB 2; Length 382;
Best Local Similarity 27.9%; Pred. NO. 2.1e-10;
Matches 95; Conservative 49; Mismatches 109; Indels 87; Gaps 17;

QY	21	IYTAIDKFNFLPVCLKIVDEDFSLPPHS--IHREIFILKTLKPHPNIIIEYFN-----DL	72
Db	40	VCAAFDTVLGINVAVKKLSRPFQNHAKRAYRELVLKCVN-HKNIISLLNVFTPQKSL	98
QY	73	KIYDDVILVTKLRYDLSQLIEITKYCKRTRFIYGINGNLVSNQYTLANEIEEKDIKLM	132
Db	99	EEFQDVYLVMLMDANLQCVIHM-----ELDHHERMSYL	131
QY	133	LKSMSSGLEFTHSQGIHRDIKPSNIFFARDDITQPIIGDFDI---CYDLKLPPKDEPP	188
Db	132	LYQMLCGIKHLHSAGIHRDLKPSNI-VVKSDCTLKIL-DFGLARTACTNFMTPY----	186
QY	189	MAKYIDVSTGIYKAPELILGITNYEYEDISLGLIILTGLYSENFQSVLVKDKKELTND	248
Db	186	-----VVTRYRAPEVILGM-GYKENVDIWSVCIM-----GELVKGCVIFQGT	229
QY	249	HVSDLYLLNQIFENFGTPNLTDFEDELFCDEYNENLHFKFKNLQKYPR-----KDWD	301
Db	230	HIDQ---WNKVIEQLGTPS-AEFMKRLQPTVRNYVE-----NRPKYPGKFEELFPDW-	279
QY	302	IILPRCND-DLMK-----EFTKMIRYDRSKRITSKEILQ	335

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: November 6, 1999, 02:20:53 ; Search time 438.25 Seconds
(without alignments)
7394.718 Million cell updates/sec

Title: US-09-072-994-13
Perfect score: 1019
Sequence: 1 ATGAAGTGTGCAGATTATTA.....TTACAATTAATGTTGGATTG 1019

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_st.*
- 14: gb_sts.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vi.*
- 18: em_fun.*
- 19: em_htg.*
- 20: em_hum1.*
- 21: em_hum2.*
- 22: em_in.*
- 23: em_om.*
- 24: em_or.*
- 25: em_ov.*
- 26: em_pat.*
- 27: em_ph.*
- 28: em_pl.*
- 29: em_ro.*
- 30: em_sts.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vi.*
- 34: gb_htgl.*
- 35: gb_htg2.*
- 36: gb_in1.*
- 37: gb_in2.*
- 38: em_bal.*
- 39: em_ba2.*
- 40: em_hum3.*
- 41: em_hum4.*
- 42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description
c 1	93.2	9.1	168935 34	HS1108D11 AL034419 Homo sapi

2	89.6	8.8	187544	35	AC004157	AC004157 Plasmodiu
3	89.6	8.8	104992	35	AC005504	AC005504 Plasmodiu
4	88.8	8.7	3542	7	YSCMTCG16	L36900 Saccharomyc
5	87.8	8.6	14867	37	AE001398	AE001398 Plasmodiu
6	86	8.4	14001	36	PFCOMP1RB	X95276 P.falciparu
7	85.8	8.4	162996	42	AC006441	AC006441 Homo sapi
8	85.6	8.4	168935	34	HS1108D11	AL034419 Homo sapi
9	85.6	8.4	187544	35	AC004157	AC004157 Plasmodiu
10	85.6	8.4	104992	35	AC005504	AC005504 Plasmodiu
11	85.6	8.4	102353	35	AC007708	AC007708 Homo sapi
12	84.6	8.3	1867	7	MTSCAJ23	AJ223323 Saccharom
13	84.4	8.3	85779	7	SCE011856	AJ011856 Saccharom
14	83.6	8.2	86829	36	PFMAL3P5	AL034556 Plasmodiu
15	83.4	8.2	2426	7	SDU49822	U49822 Saccharomyc
16	83.4	8.2	173839	35	AC007483	AC007483 Homo sapi
17	83	8.1	110000	34	PFMAL4P1_2	Continuation (3 of
18	82.6	8.1	15421	36	PFCOMP1RA	X95275 P.falciparu
19	82.4	8.1	5371	7	YSCMTCGSA	J01462 Yeast (S.ce
20	81.6	8.0	120610	34	AC004948	AC004948 Homo sapi
21	81.6	8.0	15421	36	PFCOMP1RA	X95275 P.falciparu
22	81.4	8.0	14867	37	AE001398	AE001398 Plasmodiu
23	81.2	8.0	27694	7	HASMT	D31785 Hansenula w
24	81	7.9	176552	11	AC004617	AC004617 Homo sapi
25	81	7.9	173839	35	AC007483	AC007483 Homo sapi
26	80.8	7.9	86829	36	PFMAL3P5	AL034556 Plasmodiu
27	80	7.9	4601	36	DMU11584	U11584 Drosophila
28	80	7.9	19517	37	DMU37541	U37541 Drosophila
29	79.6	7.8	12940	37	AE001386	AE001386 Plasmodiu
30	79	7.8	4601	36	DMU11584	U11584 Drosophila
31	79	7.8	16019	36	MIDYRN	X03240 Drosophila
32	79	7.8	2290	36	MIDYTRN	X05915 D. yakuba m
33	79	7.8	13594	37	AE001367	AE001367 Plasmodiu
34	79	7.8	19517	37	DMU37541	U37541 Drosophila
35	79	7.8	162996	42	AC006441	AC006441 Homo sapi
36	78.8	7.7	1200	7	MISCV	V00705 Yeast mitoc
37	78.8	7.7	2305	36	DTMTRNA	X54011 D.teissieri
38	78.2	7.7	85779	7	SCE011856	AJ011856 Saccharom
39	78.2	7.7	284972	35	AC005505	AC005505 Plasmodiu
40	78	7.7	220475	35	AC006278	AC006278 Plasmodiu
41	78	7.7	231699	35	AC006281	AC006281 Plasmodiu
42	77.8	7.6	134743	11	HUAC002300	AC002300 Homo sapi
43	77.2	7.6	27694	7	HASMT	D31785 Hansenula w
44	77.2	7.6	176552	11	AC004617	AC004617 Homo sapi
45	76.8	7.5	194410	35	AC005140	AC005140 Plasmodiu

ALIGNMENTS

RESULT	1	HS1108D11/c	LOCUS	HS1108D11	168935 bp	DNA	HTG	02-JUN-1999
DEFINITION				Homo sapiens	chromosome 20 clone 1108D11,	WORKING DRAFT SEQUENCE,		
					in unordered pieces.			
ACCESSION				AL034419				
NID				g4995265				
VERSION				AL034419.10	GI:4995265			
KEYWORDS				HTG; HTGS_PHASE1.				
SOURCE				human.				
ORGANISM				Homo sapiens				
				Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
				Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE				1 (bases 1 to 168935)				
AUTHORS				Clark,G.				
TITLE				Direct Submission				
JOURNAL				Submitted (02-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk				
COMMENT				On Jun 3, 1999 this sequence version replaced gi:4756985. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be				

contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: dj1108D11 Contig_ID: 03938 acc=AL034419 Length: 168935 bp.
* NOTE: this is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1. .168935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="1108D11"
BASE COUNT 45949 a 40608 c 39484 g 42894 t
ORIGIN

Query Match		9.1%;	Score 93.2;	DB 34;	Length 168935;
Best Local Similarity		46.5%;	Pred. No. 0.00013;		
Matches 474;		Conservative	0;	Mismatches 538;	Indels 8; Gaps 5;
QY	1	ATGAAGTTGTCAGATTATTATATAGACAAGGAATTAATTACAAATAGTCCCAATTTCTGAT	60		
Db	140252	ATATATTTATAATAACATATATTATGTATATTTTATAAATATTTATATATATTTATAAT	140193		
QY	61	ATATATACGGCTATTGATAAGTTTAATAACTTACCAGTATGCTTTA-AAATAGTTGATGA	119		
Db	140192	AAATGATATTTATGTATATTTTATAAATATTTTATATATTTTATAAATAATATCTTTA	140133		
QY	120	AGATTTCAGTCTCCACCACATTCATCCATCGAGAAATTTTATA-CTTAAACCTTGA	178		
Db	140132	TATATTTATAATAATATATATTTATAATATATTTAATGATGATATATATTTATAATATTT	140073		
QY	179	AACCACATCCAAACATAAATTGAATATTTTAAATGATCTTAAAAATTTATGATGATGTTATAT	238		
Db	140072	ATAAATATATATTTATATATATTTTATAAATAATATTTATATATTTTATAAATAATATAT	140013		
QY	239	TAGTCACCAAAATTGTATCGTTATGATTGAGTCAATTGATTGAAATTCACAAAATATTGTA	298		
Db	140012	TTATTATATTTATAATAAATAATATATTTATATATATTTATAAATAATATATTTATTTA	139953		
QY	299	AACGAACAACACCATTTATTTATGGTATTAAATGGTAATCTTGTAGTAATCAATATACAC	358		
Db	139952	TATATATTTATAAATATATATTTTATATATATATTTTATAAATAATATATTTATTTAT	139893		
QY	359	TTGCTAATGAATTGAAGAAAAGATATCAAAATTAATCGTTAAATCAATCAATGATTCAGGAC	418		
Db	139892	ATAATAAATATATATTATAAATAATATATTTATATATATTTATAAATAAT--ATATATTTA	139835		
QY	419	TGGAATTTATTCATTCACAAGGGATAAATTCATCGTGATATAAAACCCAGTAATATTTTCT	478		
Db	139834	TATATATTTATAAATAATATATTTATTTATATATATATTTTATAAATAATATATTTAT	139775		
QY	479	TGCCCCGGGATGATATAACACACACCGATTATTGGAGATTTTGGATATTTGTTATGATTTAA	538		
Db	139774	TATATATATTTATAAATAAATATATATATTTTATAAATAATATTTATTTATATATTTATA	139715		
QY	539	AACTGCCACCHAAAGATGAACCCCTATGGCGAAATATATTTGATGTATCTACAGGTATTT	598		
Db	139714	TATTTATAATAAATATTTTATTAATTTATAAATAAATATTTATATGTATATAATAAATA	139655		
QY	599	ATAAAGCACCAATTTGATCTTCGGTATAACTAATATATGAATATGAATGATTTTGGT	658		
Db	139654	TTTATTTATATGTATTTATAAATAAATATTTTATAATATTTTATAAATAATATATTTATA	139595		
QY	659	CATTGGGTATAATTTTGACTGGTTTATATTCAGAAAATTTTCAAAGTGTTCAGTCAAAG	718		
Db	139594	TATT--TATAATAAATATTTTATATATTTTATAAATAAATATTTTATATATTTATAAATA	139538		
QY	719	ATGATAAAGAATTGACTAATGATTCATGTTAGTGATTTATATTTTATAAATCAAAAT	778		
Db	139537	TTTATTTATATATTTTATAAATAAATATTTTATTTATATATTTTATAAATAAATATATTTATA	139478		

QY	779	TTGAAAAATTTCCGTACACCCCAATTTAACTGATTTTGAAGATGAATTTTGTGATGA-A	837
Db	139477	ATAAATATATATTTATATATTTTATAATAAATATATATTTATATATAAATATATTTATA	139418
QY	838	TATAATAATGAAACTGCAATTTTAAAAAATTCAAATTTACAAAAATATCCTAGAAAAAGAT	897
Db	139417	TATATTTATAATATATATATTTTATATATATTTTATAATAAATATATATTTTATATATAAT	139358
QY	898	TGGATATATTTTACCTCGATCGAATGATGATTTTAATGAAAGAAATTTTACCAGAAGTG	957
Db	139357	ATATATTTTATATATATATTTTATAATAAATATATATTTTATATATTTATAATGATATTTA	139298
QY	958	ATTAGATATGATCGAAGTAAAGATAACTTCTAAAGAAATCTTACAATTAATGTTGGAT	1017
Db	139297	TTTATATATATTTATAATAAATATATATTTTATATATTTTATAATGATATTTTATTTATATAT	139238

RESULT	2
AC004157	
LOCUS	AC004157 187544 bp DNA HTG 01-APR-1999
DEFINITION	Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 9 unordered pieces.
ACCESSION	AC004157
NID	94558582
VERSION	AC004157.3 Gi:4558582
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	malaria parasite P. falciparum.
ORGANISM	Plasmodium falciparum
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 187544) Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B. and Davis,R.W.
TITLE	Plasmodium falciparum 3D7 chromosome 12
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 187544) Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
AUTHORS	Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
TITLE	On Apr 2, 1999 this sequence version replaced gi:4337171.
JOURNAL	* NOTE: This is a 'working draft' sequence. It currently consists of 9 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

- * 47083: contig of 47083 bp in length
- * 47084 47283: gap of unknown length
- * 47284 61730: contig of 14447 bp in length
- * 61731 61930: gap of unknown length
- * 61931 80038: contig of 18108 bp in length
- * 80039 80238: gap of unknown length
- * 80239 111680: contig of 31442 bp in length
- * 111681 111880: gap of unknown length
- * 111881 142715: contig of 30835 bp in length
- * 142716 142915: gap of unknown length
- * 142916 146364: contig of 3449 bp in length
- * 146365 146564: gap of unknown length
- * 146565 151834: contig of 5270 bp in length
- * 151835 152034: gap of unknown length
- * 152035 156341: contig of 4307 bp in length
- * 156342 156541: gap of unknown length
- * 156542 187544: contig of 31003 bp in length.

FEATURES
source Location/Qualifiers
1. .187544
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"

BASE COUNT 75587 a 17670 c 18524 g 74159 t 1604 others

ORIGIN

Query Match		8.8%;	Score 89.6;	DB 35;	Length 187544;
Best Local Similarity		44.5%;	Pred. No. 0.00034;		
Matches 444;		Conservative 0;	Mismatches 549;	Indels 5;	Gaps 2;
QY	17	ATTATATAGACAAGGAATTAATTTACAAATAGTGCCATTTCTGATATATATACGGCTATTG	76		
Db	169324	ATAATAAAATATTAAATATAAAATTAATTAATAATATATAATAATAATAATAATAATTT	169383		
QY	77	ATAAGTTTAAATAACTTACCAGTATGCTTAAAAATAGTTGATGAAGATTTTCAGTCTTCCAC	136		
Db	169384	AAAAATATAAATTAATAAAAAAATAATACTAAATATTAATAAAAAATAATAATAAAT	169443		
QY	137	CACATTCAAATCCATCGAGAAAATTTTATACACTTAAAACTTTGAAACCCACATCCAAACATAA	196		
Db	169444	AAATTTTAATTAATAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAA	169503		
QY	197	TTGAATATTTTAATGATCTTAAAAATTTATGATGATGTTATATAGTCACCAAAATGTATC	256		
Db	169504	TAAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAAT	169563		
QY	257	GTATGATTTGAGTCAATTTGATTGAAATTACAAAAATATTGTAAACGAACACACGATTTA	316		
Db	169564	ATAATAATATATTTTAAACAAATTAATTAATTAATAAAATTAAGAATTTAATAATAATA	169623		
QY	317	TTTATGGTATTAATGGTAATCTTGTGTAGTAATCAATATACACTTGCTAATGAAATGAAG	376		
Db	169624	TAAATTAATTAATAAAAAATGAATTTATTTAAATAATTAATAAACAATAATATATAATA	169683		
QY	377	AAAAAGATATCAAAATTTATGGTTAAAAATCAATGAGTTCAGGACTTGAATTTATTCATTCAC	436		
Db	169684	TTATATATTTTAGTAAATAAAATAAAATTAATAATTAATAATTAATAATTTATAGTTTAAAA	169743		
QY	437	AAGGGATAATTCATCGTGATATAAAACCCAGTAATATTTTCTTTGCCGGGATGATATAA	496		
Db	169744	AATAAAAAATAAATATATAAATTAATATATAAATAAATTAATAAATAATATAAATAAATTA	169803		
QY	497	CACAACCGATTATTGGAGATTTTGATATTTGTTATGATTTTAAACTGCCACCTAAACATG	556		
Db	169804	AATATATATACAATTAATTAATTAATATATATATAATTAATTAATTAATATATATAATAA	169863		
QY	557	AACCCCTATGCGGAATATATTGTATGTATCTACAGGTATTTTATAAGCACCAGATTTGA	616		
Db	169864	TTAAATATATATATAATTAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAA	169923		
QY	617	TTCTTGGTATAACTAATATGAATATGAAAT-TGATATTTGGTCATTTGGGTATAATTTTG	675		
Db	169924	TAATTAAATTAATAATAATATAAATTAATAATATAATAATAATAATAATAATAATAATA	169983		
QY	676	ACTGGTTTATATTCAGAAAAATTTTCAAGTGTTTTAGTCAACAGATGATAAAGAATTGACT	735		
Db	169984	ATTAATAATATATTATATAAATTAATAATATATAATAATAATAATAATAATAATAATAAT	170043		
QY	736	AATGATTCATGTTAGTGATTTATATTATTAAATCAAATATTTTGAAAAATTTCCGGTACA	795		
Db	170044	AATAATATATAATAATAATAAATAAACAATAAATTAATAATTAATAATATATAATAATAAT	170103		
QY	796	CCCAATTTAACTGATTTTGAAGATGAATTAATTTGTGATGAATATAATAATAAGAACTTG	855		
Db	170104	ATAAATTAANTATATATATAATAATAATTAATTAATAATAATAATAATAATAATAATAAT	170163		
QY	856	CATTTTAAAAAATTCAAATTTACAAAAATATCCCTAGAAAAAGATTTGGGATATTAT- - -TTT	911		
Db	170164	AAATATTAACAATTAATAAAAAATATACATAAATTAATAATTAATAAAAAATTAATTAATA	170223		
QY	912	ACCTCGATGCAATGATGATTTAATGAAAGAAATTTTACCACAGATGATTAGATATGATCG	971		
Db	170224	AAATAAATTAAAAAAATTAATTAATTTTAAGATAAATATATAATAATATATATAATTAAT	170283		
QY	972	AAGTAAAGAATAAATCTCTAAAGAAATCTTACAAATTA 1009			

Db 170284 TAATAAATGTTTTATATATTAATTTAATTACACAATTAA 170321

RESULT 3

AC005504

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Apr 2, 1999 this sequence version replaced gi:4337172.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

- * 1 58642: contig of 58642 bp in length
- * 58643 58842: gap of unknown length
- * 58843 91011: contig of 32169 bp in length
- * 91012 91211: gap of unknown length
- * 91212 104992: contig of 13781 bp in length.

FEATURES

source

1..104992
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="12"

BASE COUNT 44286 a 9326 c 9564 g 41411 t 405 others
ORIGIN

Query Match 8.8%; Score 89.6; DB 35; Length 104992;
Best Local similarity 44.5%; Pred. No. 0.00041;
Matches 444; Conservative 0; Mismatches 549; Indels 5; Gaps 2;

QY 17 ATTATATAGACAAGGAATTAATTTACAATAGTGCCATTTCTGATATATATACGGCTATTG 76

Db 72789 ATAATAAAATATTAAATATAAAATTAATTAATAATATAATAATAATAATAATAATTT 72848

QY 77 ATAAGTTTAAATAACTTACCAGTATGCTTTAAATAAGTTGATGAAGATTTTCAGTCTTCCAC 136

Db 72849 AAAATATAAAATTAATAAAAAAATAATACTAATATTAATAATAAAAAATAATAATAAT 72908

QY 137 CACATTCAAATCCATCGAGAAATTTTATACCTTAAAACTTTGAAACCCACATCCAAACATAA 196

Db 72909 AAAATTTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA 72968

QY 197 TTGAATATTTTAATGATCTTAAAAATTTATGATGATGTTATATTAGTCACCAAAATTTGATC 256

Db 72969 TAAATTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 73028

QY 257 GTTATGATTTGAGTCAATTTGATTGAAATTTACAAAATATTGTAAACGAACACGATTTA 316

Db 73029 ATAATAATATATATTAAACAATTAATTTATTATAAAATTAAGAATTTATTAAATAATAATA 73088

QY	317	TTTTATGGTATTAAATGGTAATCTTGTAGTAATCAATATACACTTGCTAAATGAAATTGAAG	376
Db	73089	TAAATTAATTAATAAATAAATGAATTAATTAATAAATAAACAATAATATATATAAATAA	73148
QY	377	AAAAGATATCAAAATTTATGGTTAAAAATCAATGAGTTTCAGGACTTGAATTTATTCATTCAC	436
Db	73149	TTATATATTAGTAAATAAATAAATAAATTAATAATTAATTAATAATTTATAGTATAAAA	73208
QY	437	AAGGATAATTCATCGTGATATAAAACCCAGTAATATTTTCTTGGCCCGGATGATATAA	496
Db	73209	AATAAAATAAATATATAATTAATTAATATATAAATAAATTAATAAATAATTAATAAATA	73268
QY	497	CACACCGGATTATTGGAGATTTTGATATTTGTTATGATTTTAAACTGCCACCTAAAGATG	556
Db	73269	AATATATATACAATTAATAATTAATATATATATAATTAATTAATTAATATATAATTA	73328
QY	557	AACCCCTATGGCGAAATATATTGATGTATCTACAGGTATTTATAAAGCACCACAATTGA	616
Db	73329	TTAAATATATATATAATTAATAATTAATTAATAAATAAATAAATAAATAAATAAATA	73388
QY	617	TTCTTGGTATAACTAATATGAATATGAAT-TGATATTTGGTCATTGGGTATAATTTTG	675
Db	73389	TAATTAATAATTAATAATATATAATTAATTAATATATAAATAAATAAATAATATATA	73448
QY	676	ACTGTTTATATTCAGAAAAATTTTCAAACTGTTTGTAGTCAAGATGATAAAGAAATTGACT	735
Db	73449	ATTAATAATTATATTATATAAATAATAATAATATAATAATAAATAAATAAATAAATA	73508
QY	736	AATGATTCATGTTAGTGATTTATATTTTAAATCAATATTTTGAAAAATTCGGGTACA	795
Db	73509	AATAATATATAATATAATAAATAAACAATAATTAATAATTAATAATATATATAATAAT	73568
QY	796	CCCAATTTAACTGATTTGAAGATGAATTTTGTGATGAATATAATAATGAAAACTTG	855
Db	73569	AATTAATAATATAATAATAATAATTAATTAATAATAATAATTAATAATAATAATAAT	73628
QY	856	CATTTTAAAAAATTCAAATTTACAAAAATACCTAGAAAAAGATTTGGGATATTAT- - -	911
Db	73629	AAATATTAAACAATTAATAAATAATACATAATTAATAATTAATAAATAATTAATATA	73688
QY	912	ACCTCGATGCAATGATGATTTAATGAAGAAAAATTTTACCAGATGATTAGATATGATCG	971
Db	73689	AAATAATTAAAAAATAATTAATTTAAGATAATATATAATATATATAATAATTAATTAT	73748
QY	972	AAGTAAAGAATAACTTCTAAAGAAATCTTACAATTAA	1009
Db	73749	TAATAAATGTTTTATATTAAATTTAATTAACACAATTAA	73786
RESULT	4		
YSCMTCG16			
LOCUS	YSCMTCG16	3542 bp	DNA
DEFINITION	Saccharomyces cerevisiae mitochondrion transfer RNA-Ser1 (trRNA-Ser)		
	gene and vari gene, complete cds.		
ACCESSION	L36900	M62622	
NID	9559276		
VERSION	L36900.1	GI:559276	
KEYWORDS	transfer RNA-Ser1; vari gene.		
SEGMENT	16	of 19	
SOURCE	Mitochondrion Saccharomyces cerevisiae (organelle Mitochondrion Saccharomyces cerevisiae) DNA.		
ORGANISM	Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.		
REFERENCE	1	(bases 1 to 3542)	
AUTHORS	de Zamaroczy,M. and Bernardi,G.		
TITLE	The primary structure of the mitochondrial genome of Saccharomyces cerevisiae--a review		
JOURNAL	Gene	47 (2-3), 155-177	(1986)
MEDLINE	87163488		
FEATURES	Location/Qualifiers		
source	1.	.3542	
	/organism="Saccharomyces cerevisiae"		

/mitochondrion	
gene	/db_xref="taxon:4932"
	Join(L36887.1:3683..6340,1..1111)
trNA	/gene="trNA-Ser"
	1025..1111
gene	/gene="trNA-Ser"
	/product="transfer RNA-Ser"
	1767..2981
CDS	/gene="varI"
	/db_xref="SGD:S0007275"
	1767..2981
	/gene="varI"
	/codon_start=1
	/transl_table=3
	/protein_id="AAA67536.1"
	/db_xref="PID:g567922"
	/db_xref="GI:567922"
	/translation="MKLKLNLMLSMNKNNTNNNNIIINNTLDSLNNKKLLLNMLLD
	MNNKMMNNKRLNNNNMNPAGANPVVHRIGPAGNINNNKLNKLNMMNNWNTQIYNNK
	NMEIMNTMNDKLINKLLYKMMTLKLNMMNINKIIMSKTINQHSLNKLNIKFPYYNNNDI
	NNNNNNNNYYMMNMKLMNIMNNNNNNLNCILSYYYNNKVTIEPIKLSYIYLNSD
	IFSKYISLNDMDKYNNCILTNYQRLNNINPKLNDHNISMNYINNINNNNNKYNMMI
	NLLNNNNNNNNNNNNNNNNNNYIGNNNNNYNNMTIDNIPMDILMYKYLVGWS
	IKFKGRLSNNNGRTSTTNLLNGTFNNKKYLWSNNNNYKLNYPISNHNLYNNSNINKN
	GKYNIKVKLNFI"
BASE COUNT	1605 a 240 c 251 g 1435 t 11 others
ORIGIN	
Query Match	8.7%; Score 88.8; DB 7; Length 3542;
Best Local Similarity	45.4%; Pred. No. 0.0016;
Matches 405; Conservative	0; Mismatches 477; Indels 10; Gaps 2;
QY	14 ATTATTATATAGACAAGGAATTAATTACAATAGTGCCATTTCTGATATATATACGGCTA 73
Db	2356 ATTATTATAATAAAAAAGTAACTATTGAACCTATTAAATTATCATATATTTATTTAATA 2415
QY	74 TTGATAAGTTTAATAAAGTTTACCAGTATGTCTTAAAAATAGTTGATGAAGATTTTCAGTCTC 133
Db	2416 GTGATATTTTAGTAAAAATATATTAGTTTAAATGATATAGATAAATAATAATAATGGTATCT 2475
QY	134 CACCACATTCATCCATCGAGAAATTTTATACCTTAAACCTTTGAAACCACATCCAAACA 193
Db	2476 TAACTAATTTATCAACGTATATTAAATAATATATTATGCCTAAATTAATGATCATATATATT 2535
QY	194 TAATTGAATATTTTAATGATCTTAAAAATTTATGATGATGTTATAT-----TAGTCA 244
Db	2536 CTATAAATTTATATAATAATATTAAATAATATTAAATAATAATAATAATAATAATAATAATA 2595
QY	245 CCAAATTGTATCGTTATGATTTTGAGTCAATTTGATTGAAATTAACAAAATATTGTAAACGAA 304
Db	2596 ATTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATT 2655
QY	305 CAACACGATTTTATTTATGGTATTAAATGGTAACTCTTGTAGTAATCAATATACACTTGCTA 364
Db	2656 ATAATAATAATAATAATAATTTATTTGGTAAATTTAAATAATATTTTATAATAATAATAACTA 2715
QY	365 ATGAAATTGAAGAAAAAGATATCAAAATTATGGTTAAATCAATGAGTTTCAGGACITGAAT 424
Db	2716 TTGATAATATTCCTATAGATATTTTAAATATATAAATAATTTAGTTGGTTGATCTATTAAAT 2775
QY	425 TTATTCATTCACAAGGGATAATTCATCGTGATATAAAACCCAGTAATATTTTCTTTGCC 484
Db	2776 TTAAAGGTAGATTAAAGTAATAATAATGGTAGAAGTACACTACACTTAATTTATTAATGGTA 2835
QY	485 GGGATGATATAACACAAACCGATTTATGGAGATTTTGTATTTGTTATGATTTTAAACTGC 544
Db	2836 CTTTAAATAATAAAAAATATTTATGAAGTAATATTAATAATAATAATAATAATAATAATA 2895
QY	545 CACCTAAAGATGAACCCCTATGGCGAAATATATTGATGTATCTACAGGTATTTATAAAG 604
bb	2896 TCCCTTCTAATCATAAATTTATATAATAATCTAATATTAAATAAAAAATGGTAATAATAATA 2955

QY 605 CACCAGAAATTGATCTTGGTATAACTAAATTATGAATATGAAATTGATATTTGGTCATTGG 664
Db 2956 TTAAGTTAAATAAACTTTATTTAATATATATATTAATAGTCGGGGCCGCCACCGGA 3015
QY 665 GTATAATTTTGACTGGTTTATATTCAGAAAAATTTTCAAAGTGTTTTAGTCAAAAGATGATA 724
Db 3016 CCGGAACCCCGAAGG-AGAAATAAAATAAATAATAATAATAATAATAATAATA 3074
QY 725 AAGAATTGACTAATGATCTCATGCTAGTGATTTAGTGATTTATTAATCAAATATTGAAA 784
Db 3075 TATATATATATAAATATATAAAATAATATTTTACITTTTATATATATATATATAT 3134
QY 785 ATTTCGGTACACCAATTTAACTGATTTTGAAGATGAATATTTTGTGATGAATATAATA 844
Db 3135 AAATAAATATATATCATATAATTATATAAAATAAAATTAATTTATTTATATATAA 3194
QY 845 ATGAAAACTTGCAATTTAAAAAATTCAAATTTACAAAAATATCCTAGAAAAGA 896
Db 3195 ATATTAATTAATTTTATATAATTATTATATAATAATAATTAATTAATAAATA 3246
RESULT 5
AE001398/C
LOCUS
DEFINITION Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence.
ACCESSION AE001398 AE001362
NID 93845197
VERSION AE001398.1 GI:3845197
KEYWORDS
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 14867)
AUTHORS Gardner,M.J., Tettelin,H., Carucci,D.J., Cummings,L.M., Aravind,L., Koonin,E.V., Shalloom,S., Mason,T., Yu,K., Fujii,C., Pederson,J., Shen,K., Jing,J., Aston,C., Lai,Z., Schwartz,D.C., Pertea,M., Salzberg,S., Zhou,L., Sutton,G.G., Clayton,R., White,O., Smith,H.O., Fraser,C.M., Adams,M.D., Venter,J.C. and Hoffman,S.L.
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum
JOURNAL Science 282 (5391), 1126-1132 (1998)
MEDLINE 99021743
REMARK Erratum:[published erratum appears in Science 1998 Dec 4;282(5395):1827]]
REFERENCE 2 (bases 1 to 14867)
AUTHORS Gardner,M.J.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20814, USA
FEATURES
source
1. .14867
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
complement(1570..2424)
/gene="PFB0490c"
complement(1570..2424)
/gene="PFB0490c"
/note="predicted by GlimmerM"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC71887.1"
/db_xref="PID:93845198"
/db_xref="GI:3845198"
/translation="MKEKNEKIMDYLSCPDLDVVDREKSGKNSLLKSSSTKKSDYKK
SSIFSCKRDSHKKGSGFRGRSGFINRKSSEFKKPYNNRINKNNYKGRNFHNGR
DNFKGRTGSFGSRVFDNRKGSFKKRFISNRNKS SVKSGNSKMGKSPFNKAPTSR
TVVTKRLNNYKTVSAPVKKENNLNISLYRKNRTFALNTRKSPVGTIKSSVPRKRIKK
GLRKGSLKSTKSTSGSKFKPLNKYFLSKIKIVTSLNKPSPLEKQKNTEVNLPESL
NNATTKN"
11241..14606
/gene="PFB0495w"

CDS
11241..14606
/gene="PFB0495w"
/note="predicted by GlimmerM"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC71888.1"
/db_xref="PID:93845199"
/db_xref="GI:3845199"
/translation="MFALKKNTVREGFVNCFPSYLLKLYLKSNEFVNVLNYETNNEKR
TNKKIYKSKAQSLFDKGLNIHDKLILFKNLPKYKCAKYEICISAKEVYKYLLEYKKKC
FNYISLCDIIQSVKIFDELDTFTDYNFYIEVKNIDKNVLKINELFYFKNKDITFHR
EILGKIONKIMSYIHEMNGNELIHFLIYFFRWKNDKNLILFYNYFNYVFDHMYLFN
HEIYKLLFIFNKYLNNSNIPFNKNLIQEMEFNLYYFREIKNKNYIILKMKKEIYKK
CFKPHENVDHIDNEKILNIRLYVDSILDDIDNNKMLCNLNLNINENIEYISKLL
NFYCTLIKKGYDNDMTIYKLKEVIKATHHILCDKTKNLETFCSDIDYSTLLNSLNK
FILNKIIDKNFILFYECCLKILLNKFVNFSQSLCISLISKNIYNNILNINVYIVNV
LFNDIMKFSLYLCNIFLGRKIKTENENAVLIHNNDQTNYSNENIKDIIIOKRIKEY
IFYKMENYKDFHFKLKDSLISIKLLSNTFVKINEVNSYDFYLLFNNISCLINFLV
NRNSVKYKDTYIYILNDLSFVYKYIKNDRFKKKKNFFLLSSSMKELICKNLSVSN
RYIKHLHEEDNFDQDOYVCSLTFLNLFDFDKIIHFHYIYNLWCHVYKTYNYFKCNKL
INEDIISLLLTCSKFQFIENNSNDRYCRKELIHLKYNIIDDLKKNLYNTYKXISID
NISKFISLSNSKYTCVENLLESLOSEFEKVTKTSKGGIHMNDNNLLDNNNSCE
KYEHRVIEYKKENLFINLKIIECLIKLNIIFYLKKKTYLYLKQSLCPINLKENIL
KKILYANNLMYEMGYVCEMLERVLSSHKEONLFSYNKNVYEHKMFEDKILCHISE
DDYIEMSNMYVLFYDYLKINSESRQSNILNNSNDNRIDEIKEKKYKLNNTLIK
NNVKLYEKSNNNGNISNLKDDKNKNNHNVEMDLIDNKNENKKIQEKGQNGCEN
CKDVLNDIINIFGFLKMEKKKFLFQLYMYLCNITKFKRRYVSSSSSLFHMDFVKTIK
DMNLKYLCLENYKIKNEECALYITIDIVLEKER"
BASE COUNT 6284 a 1019 c 1106 g 6458 t
ORIGIN
Query Match 8.6%; Score 87.8; DB 37; Length 14867;
Best Local Similarity 45.0%; Pred. No. 0.0013;
Matches 448; Conservative 0; Mismatches 542; Indels 5; Gaps 3;
QY 2 TGAAGTTGTCAGATTATTATATAGACAAGGAATTAATTTACAAATAGTGCCATTTCGTGATA 61
Db 9330 TAAAAATAAATAAATAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATA 9271
QY 62 TATATACGGCTATTGATAAGTTTATAAATACTTACCAGTATGTCCTTAAAAATAGTTGATGAAG 121
Db 9270 AATTTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATTT 9211
QY 122 ATTTACAGTCTTCCACCACATTCAAATCCATCGAGAAAATTTTATACCTTAAACATTTGAAAC 181
Db 9210 AATTAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9151
QY 182 CACATCCAAACATAATTGAATATTTTAAATGATCTTAAAAATTT--ATGATGATGTATAT 238
Db 9150 TAATAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9091
QY 239 TAGTCACCAAAATTGTATCGTTATGATTTGAGTCAATTTGATTGAAATTAACAAAATATGTA 298
Db 9090 TAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 9031
QY 299 AACGAACAACACGATTTATTTATGGTATTAAATGGTAATCTTGTAGTAATCAATATACAC 358
Db 9030 ATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8971
QY 359 TTGCTAATGAAATTGAAGAAAAAGATATCAAAATATGTTTAAATCAATGAGTTCAGGAC 418
Db 8970 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8912
QY 419 TTGAATTTATTCATTCACAAAGGATAATTCATCGTGATATAAAACCCAGTAATATTTCT 478
Db 8911 ATAAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8852
QY 479 TTGCCCGGGATGATATAACACAACCGATTTATGGAGATTTTGTATATTTTGTATGATTAA 538
Db 8851 AAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 8792
QY 539 AACTGCCACCTAAAGATGAACCCCTATGCGGAAATATATTGTATGTATCTACAGGTATTT 598

repeat_region	12282. .12619
repeat_region	/rpt_family="MER1B"
repeat_region	complement(12620. .12649)
repeat_region	/rpt_family="MIR"
repeat_region	12656. .12809
repeat_region	/rpt_family="L1MA2"
repeat_region	complement(12941. .13659)
repeat_region	/rpt_family="LTR1B"
repeat_region	13662. .13779
repeat_region	/rpt_family="L1MA3"
repeat_region	13780. .14075
repeat_region	/rpt_family="AluJb"
repeat_region	14076. .14240
repeat_region	/rpt_family="L1MA3"
repeat_region	complement(14333. .14627)
repeat_region	/rpt_family="AluY"
repeat_region	14628. .14651
repeat_region	/rpt_family="AT-rich"
repeat_region	complement(14691. .14761)
repeat_region	/rpt_family="MIR"
repeat_region	15905. .16202
repeat_region	/rpt_family="AluJo"
repeat_region	complement(16333. .16724)
repeat_region	/rpt_family="MSTC"
repeat_region	complement(16777. .16993)
repeat_region	/rpt_family="MIR"
repeat_region	complement(17020. .17125)
repeat_region	/rpt_family="L1MC5"
repeat_region	complement(17141. .17428)
repeat_region	/rpt_family="AluSg1"
repeat_region	complement(17432. .17570)
repeat_region	/rpt_family="AluSg/x"
repeat_region	complement(17582. .17653)
repeat_region	/rpt_family="L1MC5"
repeat_region	17954. .18218
repeat_region	/rpt_family="AluSp"
repeat_region	18219. .18289
repeat_region	/rpt_family="(GGGA)n"
repeat_region	18359. .18405
repeat_region	/rpt_family="(TG)n"
repeat_region	18639. .18676
repeat_region	/rpt_family="(TG)n"
repeat_region	18878. .19179
repeat_region	/rpt_family="AluJb"
repeat_region	19202. .19503
repeat_region	/rpt_family="AluSx"
repeat_region	complement(20063. .20356)
repeat_region	/rpt_family="AluSx"
repeat_region	20835. .20877
repeat_region	/rpt_family="L2"
repeat_region	20878. .20952
repeat_region	/rpt_family="GA-rich"
repeat_region	complement(21223. .21533)
repeat_region	/rpt_family="AluSx"
repeat_region	complement(21991. .22245)
repeat_region	/rpt_family="L2"
repeat_region	complement(22402. .22474)
repeat_region	/rpt_family="MIR"
repeat_region	complement(22472. .22790)
repeat_region	/rpt_family="L2"
repeat_region	complement(22924. .23524)
repeat_region	/rpt_family="L2"
repeat_region	23525. .23829
repeat_region	/rpt_family="AluJb"
repeat_region	complement(23830. .23861)
repeat_region	/rpt_family="L2"
repeat_region	23936. .24244
repeat_region	/rpt_family="AluSx"
repeat_region	24247. .24443
repeat_region	/rpt_family="AluSg/x"
repeat_region	25664. .25971
repeat_region	/rpt_family="AluY"

Query Match	8.4%;	Score 85.8;	DB 42;	Length 162996;
Best Local Similarity	45.7%;	Pred. No. 0.0011;		
Matches 418;	Conservative	0;	Mismatches 487;	Indels 10; Gaps 3;
QY 33	ATTAATTTACAAATAGTCCCATTTCTGATATATATACGGCTATTGATAAGTTTAAATACTT	92		
Db 78414	ATGTTTATATATATATATATTTTATATATGTTTATATATATATATATATATATATATAT	78473		
QY 93	ACCAGTATGCTTAAATAAGTTGATGAGACATTTTCAGTCTTCCACCACATTCATCCATCG	152		
Db 78474	A--TATATGTTTATATATCTATATATATATATGTTTATATATATATATATATATATAT	78531		
QY 153	AGAAATTTTATACTTAAACATTTTGAACACACATCCAAACATAAATTGAATATTTTAAATGA	212		
Db 78532	ATATATTTGATATATTTAT	78591		
QY 213	TCTTAAATTTTATGATGATGTTTATATATAGTCACCAAAATTTGATCGTTATGATTTGAGTCA	272		
Db 78592	ATTTATATATTCATATATATTTATATATATTTATATATTCATATATATTTATATATATAT	78651		
QY 273	ATTGATTGAAATTCACAAATATTTGFAAACGAACACACGATTTTATTTATGGTATTAATGG	332		
Db 78652	ATATATTCATATATATTTATATATAGTTATATATATTCATATATATTTATATAGTTATAT	78711		
QY 333	TAATCTTGTTAGTAATCAATATACACTTGCTAATGAAATTTGAAGAAAAGATATCATCAATT	392		
Db 78712	ATATATATTTATATAGTTATATATATATTCATATATATTTATATATAGTTATATATAT	78771		
QY 393	ATGGTTAAATCAATGAGTTCAGGACTTGAATTTTATTCATTCACAAGGGATAATTCATCG	452		
Db 78772	ATATTTATATAGTTATATATATATTCATATATATTTATATAGTTATATATATTCATAT	78831		
QY 453	TGATATATAAACCCAGTAATATTTTCTTTTCCCGGATGATATAACACAAACCGATATTGG	512		
Db 78832	TTATATAGTTATATATATTCATATATATTTATATAGTTATATATATCTATATATTTAT	78891		
QY 513	AGATTTTGATATTTGTTATGATTTTAAACTGCCACCTPAAAGATGAACCCCTATGGCGAA	572		
Db 78892	ATATAGATATATATATTTATATATATATCTATATATTTATATATATTTATATATTTAT	78951		
QY 573	ATATATTGATGATCTACAGTATTTTATAAAGCACACAGAAATGATTTCTTGGTATAACTAA	632		
Db 78952	ATATTTATATATTTTATATATATTTTATATATATTTTATATATATTTTATATATTTAT	79011		
QY 633	TTATGAATATGAAATGATTTTGGTCAATTTGGGTATAATTTTGACTGGTTTATATTCAGA	692		
Db 79012	TTTATATATATTTTATACATATTTTATATATTTTATATATATTTTATATATATTTTATA	79066		
QY 693	AAATTTTCAAAGTGTTTTAGTCAAAGATGATAAAGAAATTTGACTAATGATCTCATGTTAG	752		
Db 79067	TATATTTTGTATATATTTATATATATTTTATATATATTTTATATATATTTTATATAT	79123		
QY 753	TGATTTATATTTTATTAATCAAATATTTTGAATTTTCCGGTACACCCCAATTTAACTGATT	812		
Db 79124	TTATATATTTTATGATTTTATATATTTTATATATGATATATATTTATTTATGATTTAT	79183		
QY 813	TGAAGATGAATTTTGTGATGAATATAATGAATAACTTGCATTTTAAAAAATTCAA	872		
Db 79184	ATATAAATATATATTTATGATTTTATATATAAATATATATTTATGATTTTATATATTTA	79243		
QY 873	TTTACAAAAATATCTAGAAAAGATTTGGGATATATTTTACCTCGATGCAATGATGATT	932		
Db 79244	TATATATTTTATATATATTTTATATTTTATATTTATGATTTTATATATATTTTATAT	79303		
QY 933	AATGAAAGAAATTTT 947			
Db 79304	TGATATATATATTTAT 79318			

RESULT	8
HS1108D11	
LOCUS	HS1108D11 168935 bp DNA
DEFINITION	Homo sapiens chromosome 20 clone 1108D11, WORKING DRAFT SEQUENCE
	HTG 02-JUN-1999

in unordered pieces.
AL034419
NID 94995265
VERSION AL034419.10 GI:4995265
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 168935)
AUTHORS Clark,G.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT On Jun 3, 1999 this sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: dJ1108D11 Contig_ID: 03938 acc=AL034419 Length: 168935 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1..168935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="1108D11"
BASE COUNT 45949 a 40608 c 39484 g 42894 t
ORIGIN

Query Match 8.4%; Score 85.6; DB 34; Length 168935;
Best Local Similarity 45.4%; Pred. No. 0.0011;
Matches 436; Conservative 0; Mismatches 509; Indels 15; Gaps 3;
QY 12 AGATTATTATAGACAAGGAATTAATTACAAATAGTGCCATTCTGTGATATATACGGC 71
Db 139392 ATATAAATATATATATTATAAATATATATAAATATATATTATATATAAATATATTTA 139451
QY 72 TATTGATAAGTTTAATAACTTACCAGTATGTCTTAAATAAGTTGATGAAGATTTTCAGTCT 131
Db 139452 TTATAAATATATAAATATATATTATTATATAAATATATATTATTATAAATATATAAATAA 139511
QY 132 TCCACCACATTCATCCGAGAAATTTTATACCTAAACCTTTGAAACCCACATCCAAA 191
Db 139512 ATATTATTATAAATATATAAATAAATATTATTATAAATATATAAATATTATTATATAAA 139571
QY 192 CATAATTGAATATTTTAATGATCTTAAATTTTATGATGATGTTATATTAGTCACCAAATT 251
Db 139572 TATATAAATATTTATTATAAATATATAAATATATATTATTATATAAATA----- 139620
QY 252 GTATCGTTATGATTTGAGTCAATTGATTGAAATTCACAAAATATTGTAACGAACACACAG 311
Db 139620 -TATATAAATATTTATTATAAATACATATAAATAAATATTATTATATATACATATAAAT 139678
QY 312 ATTTATTATGATTAATGGAATCTTGTAGTAATCAATATACACATGCTGCTAATGAAAT 371
Db 139679 ATTTATTATAAATATATAAATATATTATTATAAATATATA-ATAAATATATAAATAAT 139737
QY 372 TGAAGAAAAGATATCAAAATTATGGTTAAATCAATGAGTTTCAGGACCTGAATTTATTC 431
Db 139738 TTATTATAAATATATATATTATTATAAATATATAAATAAATAAATATATTATTATATAAA 139797
QY 432 TTCACAAGGGATAATTTCATCGTGATATAAAACCCAGTAATATTTTCTTTGCCGGGATGA 491
Db 139798 TATATAAATAAATAATATATTATTATTATAAATATATAAATATATATTATTATATAAATA 139857

QY 492 TATA-ACACAACCGATTATTGGAGATTTTGATATTTCTTATGATTTTAAACTGCCACCTA 550
Db 139858 TATATAAATATATATATTATATAAATATATTTATTATATATAAATAATATATTATAATA 139917
QY 551 AAGATGAACCCCTATATGGCGAAATATATTGATGTATCTACAGGTATTTTATAAGCACCAG 610
Db 139918 AATATATAAATAAATATATATTATATAAATATATAAATAAATATATTATTATATAAA 139977
QY 611 AATTGATCTCTGGTATAACTAATATGAATATGAAATGATATTTGGTCATTTGGGTATATA 670
Db 139978 TATATAAATAATATATATTATTATAAATATATAAATAAATATATTATTATAAATATATA 140037
QY 671 TTTTGACTGCTTTATATTCAGAAAATTTTCAAAGTGTCTTAGTCAAAGATGATAAAGAAT 730
Db 140038 ATATTATTATAAATATATATAAATATATATTATTATAAATATATAAATATATACATTA 140097
QY 731 TGACTAATGATCTCATGTTAGTGATTTATATTTTAAATCAAAATATTTGAAAAATTTTCG 790
Db 140098 ATATATATTATAAATATATATTATTATAAATATATAAAGATATATTATTATATAAATA 140157
QY 791 GTACACCCCAATTTAACTGATTTGAAGATGAATTTTGTGATGAATATAAATAATGAAA 850
Db 140158 TAAAAATATTATAAATATACATAAATATACATTTTATTATAAATATATAAATATTTAT 140217
QY 851 ACTTGCAATTTAAAAAATTCAAATTTACAAAAATATCTAGAAAAAGATTGGGATATATTT 910
Db 140218 AAATATACATAAATATATGTTTATTATAAATATATATATAAATATATAAATATATATA 140277
QY 911 TACCTCGATGCAATGATGATTTAATGAAGAAAATTTTACCAAGATGATFAGATATGATC 970
Db 140278 TATTAATATATATTTATATATATATAAATAAATATATATATTTATATATATATAAACC 140337

RESULT 9
AC004157/c
LOCUS AC004157 187544 bp DNA HTG 01-APR-1999
DEFINITION Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 9 unordered pieces.
ACCESSION AC004157
NID 94558582
VERSION AC004157.3 GI:4558582
KEYWORDS HTG; HTGS_PHASE1.
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 187544)
AUTHORS Hyman,R.W., Fung,E.L., Qin,F., Tamaki,T., Kurdi,O.B., Conway,A.B. and Davis,R.W.
TITLE Plasmodium falciparum 3D7 chromosome 12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187544)
AUTHORS Hyman,R.W., Qin,F., Fung,E.L., Conway,A.B. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (19-FEB-1998) Stanford DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT
On Apr 2, 1999 this sequence version replaced gi:4337171.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 47083: contig of 47083 bp in length
* 47084 47283: gap of unknown length
* 47284 61730: contig of 14447 bp in length
* 61730 61930: gap of unknown length
* 61731 80038: contig of 18108 bp in length
* 80038 80238: gap of unknown length
* 80239 111680: contig of 31442 bp in length


```
* 111681 111880: gap of unknown length
* 111881 142715: contig of 30835 bp in length
* 142716 142915: gap of unknown length
* 142916 146364: contig of 3449 bp in length
* 146365 146564: gap of unknown length
* 146565 151834: contig of 5270 bp in length
* 151835 152035: gap of unknown length
* 152035 156341: contig of 4307 bp in length
* 156342 156542: gap of unknown length
* 156542 187544: contig of 31003 bp in length.

FEATURES
    source
        Location/Qualifiers
            1..187544
                /organism="Plasmodium falciparum"
                /db_xref="taxon:5833"
                /chromosome="12"

BASE COUNT  75587 a 17670 c 18524 g 74159 t 1604 others
ORIGIN

    Query Match      8.4%;   Score 85.6;   DB 35;   Length 187544;
    Best Local Similarity 46.3%;   Pred. No. 0.0011;
    Matches 471;   Conservative 0;   Mismatches 529;   Indels 18;   Gaps 5;

QY  1  ATGAAGTTGTCAGATTATTATATAGACACGGAATTAATTTACAAATAGTGCCATTTCGTGAT 60
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170742 ATTAATTTAAATAAATAATATATTTAATAGTAATTAATATTAACAAATAATATAAATA 170683

QY  61  ATATATACGGCTATGATAAGTTTAAATAACTTACCAGTATGCTTAAATAAGTTTGATGAA 120
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170682 TTATATAATATTAATTAATTAATTAATAATAATAATTTATTTTATATATATATATATAT 170623

QY  121  GATTTCAAGTCTCCACCACACATTCAATCCATCGAGAAATTTTTFACCTTAAAACTTTTGAAA 180
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170622 TAATAATTAATTTAA--AATATAAAATTAAGAAAGATAATTTTATACCTTTTATTTAATTA 170565

QY  181  CCACATCCAAACATAAATTGAATATTTTAAATGATCTTAAATTTATGATGATGTTATATTA 240
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170564 ATATATAGTAATAAATAATTTTATGTTATTTATTAATAATAATATTTATTTATTTATTTA 170505

QY  241  GTCACCAAAATGTCATCGTTATGATTTGAGTCAATTGATTGAATTCACAAATATTTGFAAA 300
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170504 TTTATTTAAATAAATAATTAATTTTAT-----AAATATATTTTATTTTAAATTAATAATA 170450

QY  301  CGAACACACGATTTATTTATGGTATTAATGGTAATCTTGGTAGTAATCAATCAATATACACTT 360
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170449 AACATATAATTAATTAATTAATAATATATATATTTTTTTAAATAATAAATAATATATAT 170390

QY  361  GCTAATGAATTTGAAGAAAAGATATCAAAATTAATGTTTAAATCAATGAGTTCAGGACTT 420
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170389 TCACATTTTAATTAATAATAAATAAACCATTTATTTAATTAATTAATTAATAATAAATAAT 170330

QY  421  GAATTTATTCATTCACAAGGGATAATTCATCGTGATATAAAACCCAGTAATATTTCTCTTT 480
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170329 AAAATAATTTAATTTGTGAATTAATAATAATAATAAACAATTTATTAATAATTAATATATAT 170270

QY  481  GCCCGGGATGATATAACACACACCGATTTATGGAGATTTTGGATATTTTGTATGATTTAAAA 540
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170269 ----ATAATATATATATATCTTAAATTAATTAATTTTAAATTTTAAATTTTAAATAATA 170214

QY  541  CTGCCACCTAAAGATGAACCCCTATGGCGAAATATATTTGATGTATCTACAGGTATTTAT 600
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170213 ATTATTTATTAATATTAATTAATGAT----ATTTTATTAATTTGTTTAAATATTTATTTAT 170159

QY  601  AAAGCACCAGAATTGATTTCTTGGTATAACTAATTAATGAATATGAATTTGATATTTGGTCA 660
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170158 TATTTTATTAATATATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATTTAT 170099

QY  661  TTGGGTATAATTTTGACTGGTTTATATTCAGAAAATTTTCAAGTGTTTTAGTCAAGAT 720
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 170098 ATTTATAATATATTTAATTAATTAATTTGTTTATTTATTTATATATATATATATTTATTTA 170039

QY  721  GATAAAGAATTGACTAATGATTCCTCATGTTAGTGATTTATATTTTATTAATCAAAATATTT 780
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
```

```
Db 170038 ATTATTATTTTATTATTATATATATATTTATTAATTTATATAATAATAATTAATATATAT 169979

QY  781  GAAAATTTCCGTACACCCAAATTTAACTGATTTTGAAGATGAATTAATTTTGTGATGAATAT 840
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 169978 ATATATTTAATTTATTTATATATATTAATTTAATTTATATATATTTTATTTAATTTATATAT 169919

QY  841  AATAATGAAAACCTTGCAATTTTAAAAAATTCAAATTTACAAAAATATCCTAGAAAAAGATTGG 900
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 169918 TTTATTTATTAATTT--TATTTTATTTATTAATTTAATTTAATTTATATATATATTTAATTT 169861

QY  901  GATATTATTTTACCTCGATGCAATGATGATTTAATGAAGAAATTTTACCACAGATGATT 960
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 169860 AATTATATATATATTTAATTTAATTTATATATATATATTTAATTTAATTTGATATATATTTAA 169801

QY  961  AGATATGATCGAAGTAAAGAAATAACTTCTTAAAGAAATCTTACAAATTAATGTTGGATT 1018
    |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 169800 TTTATTTATATATTTTATTAATTTTATTAATTTAATTTAATTTATATATATTTATTTTATTT 169743
```

```
RESULT 10
AC005504/c
LOCUS
DEFINITION
    AC005504 104992 bp DNA HTG 01-APR-1999
    Plasmodium falciparum chromosome 12, WORKING DRAFT SEQUENCE, 3
    unordered pieces.
ACCESSION
    AC005504
    94558584
VERSION
    AC005504.3 GI:4558584
KEYWORDS
    HTG; HTGS_PHASE1.
SOURCE
    malaria parasite P. falciparum.
    Plasmodium falciparum
    Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE
    1 (bases 1 to 104992)
    HYMAN,R.W., FUNG,E.L., QIN,F., TAMAKI,T., KURDI,O.B., CONWAY,A.B.
    and DAVIS,R.W.
    Plasmodium falciparum 3D7 chromosome 12
    Unpublished
REFERENCE
    2 (bases 1 to 104992)
    HYMAN,R.W., QIN,F., FUNG,E.L., CONWAY,A.B. and DAVIS,R.W.
    Direct Submission
    Submitted (21-AUG-1998) Stanford DNA Sequencing and Technology
    Center, Stanford University, 855 California Avenue, Palo Alto, CA
    94304, USA
COMMENT
    On Apr 2, 1999 this sequence version replaced gi:4337172.
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 3 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    * 1 58642: contig of 58642 bp in length
    * 58643 58842: gap of unknown length
    * 58843 91011: contig of 32169 bp in length
    * 91012 91211: gap of unknown length
    * 91212 104992: contig of 13781 bp in length.
FEATURES
    Location/Qualifiers
        source
            1..104992
                /organism="Plasmodium falciparum"
                /db_xref="taxon:5833"
                /chromosome="12"
BASE COUNT  44286 a 9326 c 9564 g 41411 t 405 others
ORIGIN
```

```
Query Match      8.4%;   Score 85.6;   DB 35;   Length 104992;
Best Local Similarity 46.3%;   Pred. No. 0.0013;
Matches 471;   Conservative 0;   Mismatches 529;   Indels 18;   Gaps 5;

QY  1  ATGAAGTTGTCAGATTATTATATAGACACGGAATTTAATTTACAAATAGTGCCATTTCGTGAT 60
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 74207 ATTAATTTAATAATTAATAATATATTTAATAGTAATTAATAATTAACAAATAATAATAATA 74148
```

QY	61	ATATATACGGCTATTGATAAGTTTAATAA	CTTACCAGTATGCTTTAAAAATAGTTGATGAA	120
Db	74147	TTATATAAATATTAATAATTAATAAATAA	TAATTTATTTTATAAATTATATATATATAT	74088
QY	121	GATTCAGTCTCCACCACATTC	CAATCCATCGAGAAATTTTATACCTTAAAACCTTGAAA	180
Db	74087	TAATAAATAATTAA--AATATAAATAA	GAGATAAATTTATACCTTTTATTTAATTAATAA	74030
QY	181	CCACATCCAAACATAAATGGAATATTT	TAATGATCTTAAAATTTATGATGATGTTATATTA	240
Db	74029	ATATATAGTAATAAATAATTTTA	GTATTTATATAAATAATTTATTTATTTATTTA	73970
QY	241	GTACCAAAATGTATCGTTATGATTTGAGT	CAATTGATTCGAAATTCACAAAATATGTAAA	300
Db	73969	TTTATTTAATAAATAAATAATTTTA	ATTTTAAATATTTTAAATTAATAATAATA	73915
QY	301	CGAACAAACAGGATTTATTTATGGTATTA	ATGTTAGTAACTTTGTAGTAATCAATATACACTT	360
Db	73914	AACATATAAATTAATAATAATTAATA	TATTTTTTAAATATAATAATAATAATAAT	73855
QY	361	GCTAATGAAATGAAGAAAGATATCAAAAT	TATGGTTAAAATCAATGAGTTCAGGACTT	420
Db	73854	TCACATTTTAATTAATAAATAAACCATT	TATTAATTAACCTTAATTAATATATAAATAA	73795
QY	421	GAATTTATTCATTCACAAGGGATAATTC	ATCATCGTGATATAAAACCCAGTAATATTTCTTT	480
Db	73794	AAAATAATTTAATGTGTAATTAATAATA	TAAACATTTATTAATAATTAATATAT	73735
QY	481	GCCCGGGATGATATAACACAACCGATT	TATTTGGAGATTTTGTATATTTGTTATGATTTAAAA	540
Db	73734	---ATAATATATATATATCTTAAATTA	ATTTTAAATTAATTTTAAATATAATA	73679
QY	541	CTGCCACCTAAAGATGAACCCCTATGCG	GAAATATATTGTATCTACAGGTATTTAT	600
Db	73678	ATTATTTAATAATTAATTAATGTA	-----ATTTTAAATGTTTAAATATTTTAT	73624
QY	601	AAAGCACAGAAATGATCTTGGTATAACT	AAATATGAATATGAAATGATATTTTGGTCA	660
Db	73623	TATTTTATTTAATATATTAATAATTA	ATTAATTAATTAATTAATTAATTAAT	73564
QY	661	TTGGGTATAATTTGACTGGTTTATATTC	CAGAAAATTTTCAAAGTGTTTTAGTCAAAGAT	720
Db	73563	ATTATAATATATTTAATTTAATTAAT	TATTTGTTTATTTATATATATATATATATTTATTTA	73504
QY	721	GATAAGAAATTGACTAATGATTTCTCAT	GTTAGTGATTTATATTTTAAATCAAATATTT	780
Db	73503	ATTATATTTTATTTATATATATATAT	TATTAATTAATTTATATAAGATAAATTTAATTTAT	73444
QY	781	GAAAATTTCCGTACACCCAATTTAACT	GATTTTGAAGATGAATTTTGTGATGAATAT	840
Db	73443	ATATAATTAATTTATATATATTAAT	TTAATTTATATATTTTAAATTTAATTTATAT	73384
QY	841	AATAATGAAAACCTTGCAATTTAAAAA	ATTCAAATTTACAAAAATATCCTAGAAAAGATTGG	900
Db	73383	TTTATTTATTAAT--TATTTTATTTA	ATTAATTTAATTTAATTTATATATATATATTTT	73326
QY	901	GATATTTATTTTACCTCGATGCAATGAT	TTTAAATGAAGAAAAATTTTACCAGATGATT	960
Db	73325	AATTATATATATATTTAATTTAATTTA	TATATATATATATTTAATTTAATTTGATATATATTTAA	73266
QY	961	AGATATGATCGAAGTAAAGAAATAACT	TTCTAAAGAAATCTTACAAATTAATGTTGGATT	1018
Db	73265	TTTATTTATATATTTTATTTAATTTA	TTTATATATATTTAATTTAATTTATATATTTTATTTT	73208

RESULT	11
AC007708/C	
LOCUS	AC007708 102353 bp DNA
DEFINITION	Homo sapiens, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION	AC007708
NID	95032328
VERSION	AC007708.2 GI:5032328
KEYWORDS	HTG; HTGS_PHASE1.

```

SOURCE
ORGANISM      Homo sapiens
               Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 102353).
AUTHORS      Zhan,M. and Roe,B.A.
TITLE        Homo sapiens Chromosome 22q11 BAC Clone b563b9 In BCRL2-GGT Region
JOURNAL      Unpublished
REFERENCE     2 (bases 1 to 102353)
AUTHORS      Zhan,M. and Roe,B.A.
TITLE        Direct Submission
JOURNAL      Submitted (02-JUN-1999) Department Of Chemistry And Biochemistry,
               The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
               OK 73019, USA
COMMENT      On Jun 10, 1999 this sequence version replaced gi:4966391.
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 5 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *
               * 1 7324: contig of 7324 bp in length
               * 7325 7399: gap of unknown length
               * 7400 31342: contig of 23943 bp in length
               * 31343 31417: gap of unknown length
               * 31418 57661: contig of 26244 bp in length
               * 57662 57736: gap of unknown length
               * 57737 78241: contig of 20505 bp in length
               * 78242 78316: gap of unknown length
               * 78317 102353: contig of 24037 bp in length.
               *
               * Location/Qualifiers
               * 1..102353
               * /organism="Homo sapiens"
               * /db_xref="taxon:9606"
EASE COUNT   28695 a 23538 c 22458 g 27362 t 300 others
ORIGIN

```

Query Match 8.4%; Score 85.6; DB 35; Length 102353;
Best Local Similarity 46.0%; Pred. No. 0.0013;
Matches 396; Conservative 0; Mismatches 459; Indels 5; Gaps 3;

QY	153	AGAAATTTTATACCTTAAACCTTTGAAACCCACATCCAAACATAAATTGAATAATTTTAATGA	212
Db	57442	ATATATATATATATTTATCTGTAATATATAATTATATATAAATTATACAAATTATATAT	57383
QY	213	TCCTAAAATTTATGATGTTATATATTAGTCACCAAAATTGTATCGTTATGATTGAGTCA	272
Db	57382	AATTATATAAATTATCCAAATTCACATATAATTATATATATATAT-TATATGTAATAATTAT	57324
QY	273	ATTGATTGAAATTACAAAATATTGTAACGAACAACCGATTATTATTATGGTATTAATGG	332
Db	57323	ATATAATTATATATGATATTATATGTAATAAATTATATGTAATTATATATGATATATATG	57264
QY	333	TAAT--CTGTGTAGTAATCAATATACACTTGCCTAATGAAATTGAAGAAAAGATATCAA	389
Db	57263	TAATAAAATATATAATTATATATGATATTATATATGTAATAAATTATATATAATTATATATGAT	57204
QY	390	ATTATGGTTAAAATCAATGAGTTCAGGACTTGAATTTATTCATTCACAAGGGATAATTCA	449
Db	57203	ATTATATGTAATAATTATATATAAATTATATATGATATTATATGTAATAAATTATATATAAG	57147
QY	450	TCGTGATATAAAACCCAGTAATATTTCTTTGCCCGGGATGATATAACACAAACCGATTAT	509
Db	57143	TATATATGATATTATATGTAATAAATTATATATAAGTATATATATTATATTATATGTAATAAT	57084
QY	510	TGGAGATTTTGTATTTGTTATGATTTAAAACTGCCACCTAAAGATGAACCCCTATGGC	569
Db	57083	TATATATAAGTATATATTATTTATGTAATAAATTATATATAAGTATATATTATATTATAT	57024
QY	570	GAAATATATTGATGTATCTACAGGTATT-TATAAAGCCACCAGAATTGATTTCTTGGTATAA	628

Db 964 TTATAAAATTATATATATATATATATAATAAATAGTTCTTCTTACATAAAATATA 905

QY 980 GAATAACTTCTAAAGAAATCTTACAAATTAAT 1010

Db 904 TAAAAATTATAAAATATATTAAATTATTAAT 874

RESULT 13

SCE011856

LOCUS SCE011856 85779 bp DNA PLN 19-JAN-1999

DEFINITION Saccharomyces cerevisiae complete mitochondrial genome.

ACCESSION AJ011856

NID 94160362

VERSION AJ011856.1 GI:4160362

KEYWORDS

SOURCE baker's yeast.

ORGANISM Mitochondrion Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 85779)

AUTHORS Foury,F., Roganti,T., Lecrenier,N. and Purnelle,B.

TITLE The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae

JOURNAL FEBS Lett. 440 (3), 325-331 (1998)

MEDLINE 99087401

REFERENCE 2 (bases 1 to 85779)

AUTHORS MIPS.

TITLE Direct Submission

JOURNAL Submitted (16-DEC-1998) Data collected by MIPS on behalf of the European yeast mitochondrial genome sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, Germany; E-mail: Mewes@mips.biochem.mpg.de

FEATURES

Source

1. 85779

/organism="Saccharomyces cerevisiae"

/mitochondrion

/strain="FY1679, isogenic derivative of S288C"

/db_xref="taxon:4932"

1. .11

/note="9S ribosomal RNA"

731. .802

/gene="tp(TGG)Q - systematic name"

/note="tRNA-Pro - common name; anticodon gene: TGG"

731. .802

/gene="tp(TGG)Q - systematic name"

3952. .4338

/note="ORF Q0010"

/codon_start=1

/transl_table=3

/protein_id="CAA09821.1"

/db_xref="PID:e1370687"

/db_xref="PID:g4160363"

/db_xref="GI:4160363"

/translation="MYIMFLYNMLLIILIFYSIGVPIIIFNNYYWDPDIFLFI

YFIKFIILFNLYLYMINYIVYTPSGSPGRTYILLYNNLYSYNMFIDYVMKFITC

VTYMLMFLLSPTSPYVSEVPVS"

complement(4012. .4312)

/note="ori1"

4254. .4415

/note="ORF Q0017"

/codon_start=1

/transl_table=3

/protein_id="CAA09822.1"

/db_xref="PID:e1370688"

/db_xref="PID:g4160364"

/db_xref="GI:4160364"

/translation="MCATYMFNITVIITHPTTLRTRGPFVRNRDLIYKYKSNLIN

NLNNMTYIL"

6546. .8194

/note="15S ribosomal RNA"

9374. .9447

/gene="tw(TCA)Q - systematic name"

/note="tRNA-Trp1 - common name; anticodon gene: TCA"

9374. .9447

/gene="tw(TCA)Q - systematic name"

11667. .11957

/note="ORF Q0032"

/codon_start=1

/transl_table=3

/protein_id="CAA09823.1"

/db_xref="PID:e1370689"

/db_xref="PID:g4160365"

/db_xref="GI:4160365"

/translation="MLMMYMLFIMMKTYPMLSYHMSYHHIMLYTIMWYMKYSTYMLW

LLYKSYFIFIFITNNYNNYNYWYVTLMTNTLYLYNNMIHFLTINKKFLYSL"

complement(12510. .12780)

/note="ori8"

13818. .26701

/gene="COX1"

13818. .23167

/gene="SCA15alpha"

13818. .21935

/gene="I-SCEII"

13818. .19996

/gene="I-SCEIII"

13818. .18830

/gene="SCA12"

13818. .16322

/gene="SCA11"

/db_xref="SGD:S0007261"

join(13818. .13986,16435. .16470,18954. .18991,20508. .20984,

21995. .23167)

/gene="SCA15alpha"

/note="ORF Q0070"

/codon_start=1

/transl_table=3

/protein_id="CAA09829.1"

/db_xref="PID:e1370695"

/db_xref="PID:g4160371"

/db_xref="GI:4160371"

/translation="MVQRWLYSTNAKDIAVLYFMLAIFSGMAGTAMSLIIRLELAAPG

SOYLHNSQLFNLVVGHAFLMIFFLVMPALIGGFGNYLLPLMIGATDTAFPRINNIA

FWLPMGLVCLVTSLVESGAGTGWTVPPLSSIOAHSGPSVDLAIFALHLTSISLL

GAINEFIVITLNMRTNGTMHKLPLFVWSIFITAFLLLSLPVLSAGITMLLLDRNFT

SFFEVSGGDPILYEHLPWFHFFGHEVYILIPGFGIISHVYSTYKKPVGEISAVYA

MASIGLLGLVWSHMYIVGLDADTRAYTSATMIIAIPGKIFSWLMNPFSDKNK

MNKKLIIRYQKNNNNMKTYLNNNNMMNMKYKGNLYDIYPRSNRYIQPNINKE

LKVYGNLESCVGMPTYNIVKHMVGIPNNILYIMTGILLTDGWDYTSKDLDRKTI

MEINCRFLKQSMHSEYLMYVFMLLSHYCMSYPMKIAKVKGKSYNQLEFYTRSLPC

FTILRYMFYNGRVKIVPNLDLLNYESLAHIMICDGSFVKGGLYLNLOSFTRKELI

FIMNLIKFNLCNLTLLHKSRRKTYIYMRVESVKRFLFPMIYKILPSMRYKFDIMLWQK

KYNMIN"

13818. .16322

/gene="SCA11"

/note="ORF Q0050"

/codon_start=1

/db_xref="SGD:S0007261"

/transl_table=3

/protein_id="CAA09825.1"

/db_xref="PID:e1370691"

/db_xref="PID:g4160367"

/db_xref="GI:4160367"

/translation="MVQRWLYSTNAKDIAVLYFMLAIFSGMAGTAMSLIIRLELAAPG

SOYLHNSQLFNGAPTAYIISLMRTALVLIINRYLKHMTNSVGANFTGTWACHKTPM

ISVGVKCYMVRLTNFLQVIRITISSYHLDVVKQVLFYVEVIRLWFIVLDSTGSVK

KMKDINTKGTKSEGSTERNSGVDRGMVVPNTOMKMRFLNQVRYYSVNNNLKMGKD

TNIELSKDTSDLLFEKLVMDNMNEENNNLLSIMKNVDMLMLAYNRKSKPGNM

TPGTLETLDGMMMYLNKLSNELGTGKFKFKPMRMVNIIPKGGMRPLSVGNPRDKI

VQEVNRMILDTIFDKRMSTHSHGFRKNMSCQTAIWEVRNMEGGSNWFIEVDLKKCFDT

ISHDLIIKELKRYISDKGFIDLVIYKLLRAGYIDKGYHKPMLGLPQGLISPLICNI

VMTLVDNWLEDYINLYNKGKVKQHPYTKLSRIAKMFKSTLKLHKKERAKGPTFI

YNDPNFKRMKYVRYADDILIGVLGSKNDCKMIKRDLLNNFLNSGLTMNEEKTLLTCAT

ETPARFLGYNISITPLKRMPTVTKTIRGKTIRSRNTTRPIINAPRIINKLATNGYC

KHNKNGRMGVPTRVGRWYEEPTIINNYKALGRGILNYKILATNYKRLRERYVLY

YSCVLTASKYRLKTMSTIKKFGYNLNIENDKLIANFPRNTFDNIKKIENHGMFY

see http://www.sanger.ac.uk/Projects/P_falciiparum. IMPORTANT: This sequence is not the entire insert of clone MAL3p5. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

FEATURES

source

Location/Qualifiers

```
1..86829
/organism="plasmodium falciparum"
/strain="3D7"
/db_xref="taxon:5833"
/chromosome="3"
/clone="MAL3p5"
324..2944
/gene="MAL3p5.1"
join(324..668,1199..1303,1460..2944)
/gene="MAL3p5.1"
/note="predicted using hexExon; MAL3p5.1 (PFC0575w),
Hypothetical protein, len: 645 aa"
/codon_start=1
/protein_id="CAB38969.1"
/db_xref="PID:e1407789"
/db_xref="PID:g4493933"
/db_xref="GI:4493933"
/translation="MYLKNVYIYISSCFILFDLCFSFHLKMKYKNHMMNMKSVTFPL
RSPQIYKRFRKRRIKNSVFKKQKPLFLENLKGFSFLGFRNQYDOVNRKREKK
KKKKKKKKPKVHSILNQISEVKERKDAENLALHLFLKLDENITLFSMMHIMDF
FKSKQVIECDIKSKKKRKNLSIYNLFICTLIYFTYCMCLLIKYISHLCIFFFF
FFCFFLCYNILERIYECCVGLIRKKIERYNLYCEKKIKFHKMDAIKKMEINMKDDD
LYFNHYDELLRCFTMKLNIERNKNIIRSNDNIINNDISIDKDMYNNPIDVNNII
SLDEKIKEQFENDDENLKLKDTYEQQLFNDNLIKIYEEDOPLYNINDSNINDNN
IEQADIEFQNFSLNVNLDQGRVKSNDENTKSTEHKKNKTINKGYDTELIQNMENN
FIKKNIDNNISDISNNEINIILKKLNOSEDENLTSOLIYERLRTKVLWYIQKIEY
LKFKYQYDIINEQPIIKNEKTVLDLLNIGYKIVMSPDVDNSLFEKTKIDSIPNEKDK
NNQMENQKSNKIIMFNVN"
```

gene

CDS

```
complement(3356..6646)
/gene="MAL3p5.2"
complement(3356..6646)
/gene="MAL3p5.2"
/note="predicted using hexExon; MAL3p5.2 (PFC0580c),
Hypothetical protein, len: 1097 aa"
/codon_start=1
/protein_id="CAB38968.1"
/db_xref="PID:e1407788"
/db_xref="PID:g4493932"
/db_xref="GI:4493932"
/translation="MLGLKRKNVYLLVSPVSLFAYFLKRHKDNENNYETLINNDIE
KIKKIRIHNKCSYIPLILNIYDSYIYKNIKILRWLYFKFRKRKDKEEYIYITNMVRK
KRREAIAKYNFISDEQNLNFKFYIYEIVLEYSUKYGILSPHLSLYILKNISEHCVNIYP
SLYYNKLDNKHNLINEKCLKYFKQINNEHTQOAPTHTTHNNNNNNKKPLDINIHC
KNTNISSYTYNNMEKENIYDKNIHNFYTEKSISYKDENCQHITLNMIIYLLNQTY
DNICRICLNTNTNIYINFYMINILKYICYKNMEIILLNHNHIEDMKKIKNQKNNTIS
LFKYIYFFFFKKEENHIYDLFEDQNMNHLHKENDKFYNSNENTHNNIYKIYISDNY
FYDHINSSNRCSFKNLKKQOTDDNTKHIIMGKEKYPMNKSDHEKKNNTTCGNINIEK
DOKKDILKKIYFLKGNKLDIDQILNELYMIYMRLLFECSLKLISIKKNHLLLEKME
FDKDNKIIYLSADYMNLRNILKRFSKNEERENINSFASFPLLKSNIIYFEDEIG
RSRDNTIYNNVYDKETNKTNTNNNDNDNICSNNDHICSNNNDHICSNNNDHICSN
NNNICSNNNNNICSNNNNICSNNNNICSNMKLDDEFCQDNKFNNDYTRKKEKRRI
YELAKIYTNNIFDYLKGKKEKHQNEENTINLYIKKFPWIFYLKNIIKNKDTSFIEH
NNIVNGDINKNNIIFKKYNLFESSIISFYIKDIYEYNIYFKLRLYIYDNLIKFKCR
YFLKMNHEHINRKLKMKRAPHYIYINFDOFIINNYYHIIHKKNIHKIHHLKQCKDKE
IDIVFKDLYCYMINNINIFSIIHKVDHNECVRFKAYNKILLIYEYNLNEKENIY
YKKNIKKYLTYLNNNISNDLYPYNISYKNYQNKYKRNKNFSHIFSLKNDIDHLLLF
LYTQRIQNCDDIFSYIYIKYNFNEKNPFLNLYELHYIVYSEKKKKKFFSFISSSP
YSYDTMVSNTFSYFFFSLSYLLFILFYHPDMYASYIFFEKTTLTYSGLPTYYSLYNNI
MVVCGPKT"
```

gene

CDS

```
complement(13655..19355)
/gene="MAL3p5.4"
complement(join(13655..17801,18019..19355))
/gene="MAL3p5.4"
/note="predicted using hexExon; MAL3p5.4 (PFC0590c),
Hypothetical protein, len: 1828 aa"
/codon_start=1
```

```
/protein_id="CAB38986.1"
/db_xref="PID:e1407806"
/db_xref="PID:g4493950"
/db_xref="GI:4493950"
/translation="MLFYKLIQLFLSFIIIIHGHKDISSTKLLKSGNLNNKIKKGRRE
RYGGIYNSKYMTIEKRKKYNYIKKQKNNNSYFSCYVYKNNDVNNNYTAYNIYINNP
KNKLKEYYEKIKNHVIKKKKKIFSLKFSQNKRNKKKYYFFINFTSFHDIKNDIKVLFF
NIDYTENNIISYIKSKFRTPPTIKIYLLGTFLLSVLHMKNVYKLIILFDENKIFPKK
GEIWRLETPYLYIGNLYQYILMFNVLNIYMSVSSVEISHYKKPEDFLIFLTFGYISNLL
FTIWMYNNENIMNVKLYIHNFKNFFIKDCVSKYTSRSTNNNSNNINSSNNNN
HYNNSKNIDIKKEQYNHLGYVFSTVILYWSRINEGTLINCFELFFIKAEYVPPFFII
QNILLYNEFSLYEVASIFSSYLFTEYKYFKFNYLRLFPFKTLKVIHIYPLYDRFQMN
TKLLISLKKRKNGLPFLQCYIHNVPRINTMKHMNISDLRKNIEVCNKKNVKHKNV
WSNFLYIILKLFNKEIKNYVDFMILKLLSKYIKIEKVVLLYICEQIEHEIYKFRTR
DLTLLILKKNFNDIYYINLISKILMKMNKMSYKDALIYYLSKNIYLTDEQI
YNKEIFNFSILKFENHLNVNINLHSLSLFFYSYSVYFINCFYIYYFHSFFNIITK
FINIINKNLHLYNSTDLMFYIGSLHIHNMYPHVDONKEPKNNQKNNYHNDNHN
IYLNKINNNCYDHRLDSNDFITMNYDOGEYNKHIQONKHIQONKHIQONKH
IQIRIGTHCTESNNOQLIQIONDEKENKLITYDNSKHNLLKDPCHNIVEKDGEKKQ
NLIKNIILNIKKIIEEKLSSFKIQEIVNILEVSLNKNIIINKKYFHFNLQEKINIRNY
INIYVNINKIYLNDEEENTSHCILKIKNDNKKDILYHDMKFLYLNLMNEIYYRNDLLN
MKQIILLYGLKFNFMFLOFEKIIILKRFICLPKKEIQKIGKEEIMFLYQYFFVRTCL
FNELKQNNLFIQSDEYENYIYISDKYNESAKLDNSYNMPSNLKEKNTNHHGKDNL
DLYIHDDLFLYMNKNKRDYKIYLYDNFIFENYPAYYVEQKDHIDYNESVNNFDMKS
FIOLKKKIKINNNNNNNNNNIYIDTNIQTVNKNYSCTHNNVIKNETNDNYPNS
TIRNQHNDQVILLNPPVFFYNKLVVDSDIDFEYELTCYNLYLDIYKIVCLKLLTLK
NHKLSQSDIDILCIYEKLNIRDYRIKLYLNKLELLYLDNTYLLKVINIIVKFNLY
NMISYLIQINKLITFIYNNINNESIQILKLIQMLISVHKHNKLSPFHMMNLNVQNAAY
LFKNLYNLQNTQDLKFIEMNVYDNLTFKFKLFKNILSINVKRYVQNCNSYKEMN
THTNNLNKNEOHKYIHHNNDHKDGRHNNNNHHYDKVDVSSSSSSSYYYLKNSGKNLG
NINVQNDDININKISISYKIKKQIKDIGYMRVSKYSELKMSKMMNMYDEHFNDEY
RNVCEIYEDFLIYNKNIQVYKNINICNYTFPMAINLLTLNDENILINIKSDDNK
KLIVDKKKFLIVDILYNDYIYTLTKSKLDKLEINYIYLSYVSHIKKKKKILNYK
KYALLKLIKRGFNICYDADTYVKNKKGKSKDLSYEINKLYINNLILDILKROKKNH
LHPHPHTQNTKQIKNINIKNKLILLYHQNKKNVKKIIHFKNYKYKIMNLPDQRNHYH
NKRIKIIKDSLLAINHKTNIIEKQIKISTSNHLSKLRMFSL"
complement(20528..21454)
/gene="MAL3p5.5"
complement(20528..21454)
/gene="MAL3p5.5"
/note="predicted using hexExon; MAL3p5.5 (PFC0595c),
Serine/threonine protein phosphatase (PP2), len: 309 aa;
Similarity to serine/threonine protein phosphatases.
M.domestica serine/threonine protein phosphatase
(TR:Q42912) BLAST Score: 1005, sum P(1) = 6.9e-107; 60%
identity in 301 aa overlap."
/codon_start=1
/protein_id="CAB38970.1"
/db_xref="PID:e1407790"
/db_xref="PID:g4493934"
/db_xref="GI:4493934"
/translation="MAKEERKWIQLRMNPPKLLDESLRLVCQRKEILLVEENNVO
SIKPPVVICGDIHGQFFDLLELFDVGGDIMMNDYIFLGDYVDRGYSVETFEYLLLLK
LLFPKNITLLRGNHESROITTVYGFYDECFKYGANAWKYCTDIFDYLTLAALVDNQ
IFCVHGGLSPEIKLIDQLRLINRVQEIPEHGAFGDIMWSDPDVEDDWWANPRGAGWLF
GPNVTKKFNHINLLEIARAHQLAMEGYRYMFEDSTIITVWSAPNTCYRCGNVAAMR
IDEYMNROMLIFKDTPDSRSIKNKATIPYFL"
25252..26157
/gene="MAL3p5.6"
join(25252..25296,25453..26157)
/gene="MAL3p5.6"
/note="predicted using hexExon; MAL3p5.6 (PFC0600w),
Hypothetical protein, len: 250 aa"
/codon_start=1
/protein_id="CAB38972.1"
/db_xref="PID:e1407792"
/db_xref="PID:g4493936"
/db_xref="GI:4493936"
/translation="MKKYLKNKYMIYIYNIYNKLEEKYKNFLKKNMNSHMGASQNMNVN
NNYTMELEEFEEKINNNYNNNNNNNNINNNYNNYDMNIIKVSQSQVQHNKRLQDFYNNKN
SFQHYIKKLTCTCFDADDIRNLLLEKRLAYERDNTLIKNIQEEENKKGIGINGFGSES
NSSSSNYDNNYLLYKINRLNKTNTNKSKNRSRKRKRINSIDKKYIIKRACKFIPN
NGFKIEDYITCQNCGYNDFSIVIRSTSPNNAD"
```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 6, 1999, 02:12:43 ; Search time 288.18 Seconds
(without alignments)
6974.848 Million cell updates/sec

Title: US-09-072-994-13
Perfect score: 1019
Sequence: 1 ATGAAGTTGTCAGATTATTA.....TTACAATTAATGTTGATTG 1019

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	54	5.3	610	44	AU038896	AU038896 AU038896
2	53.8	5.3	519	40	C91042	C91042 C91042 Dict
C 3	53.6	5.3	488	45	AU000986	AU000986 AU000986
4	53.4	5.2	705	43	AI238819	AI238819 GH14939.5
C 5	52.6	5.2	1145	48	AI581441	AI581441 SWOV3MCAM
6	52.4	5.1	684	44	AI294319	AI294319 LP07659.5
7	52.4	5.1	590	44	AU037557	AU037557 AU037557
8	52.2	5.1	478	43	AU033909	AU033909 AU033909
9	52	5.1	424	40	C92817	C92817 C92817 Dict
10	51.8	5.1	755	35	AA550575	AA550575 1740m3 gm
11	51.8	5.1	712	49	AU056862	AU056862 AU056862
12	51.6	5.1	579	44	AU038989	AU038989 AU038989
13	51.4	5.0	717	45	AU000694	AU000694 AU000694
14	51.4	5.0	688	45	AU001063	AU001063 AU001063
15	51.4	5.0	593	45	AU002045	AU002045 AU002045
16	51.4	5.0	633	45	AU004482	AU004482 AU004482
17	51.4	5.0	700	45	AU006427	AU006427 AU006427
C 18	51.2	5.0	717	45	AU000694	AU000694 AU000694
C 19	51	5.0	540	35	AA550445	AA550445 1598m3 gm
C 20	50.8	5.0	450	49	AI629364	AI629364 fc05e12.y
21	50.6	5.0	539	38	AA440967	AA440967 LD12786.5
22	50.4	4.9	594	38	AA539867	AA539867 LD19149.5
C 23	50.4	4.9	688	45	AU001063	AU001063 AU001063
24	50.4	4.9	736	45	AU001091	AU001091 AU001091
C 25	50.4	4.9	593	45	AU002045	AU002045 AU002045
C 26	50.4	4.9	633	45	AU004482	AU004482 AU004482
C 27	50.4	4.9	700	45	AU006427	AU006427 AU006427
C 28	50.2	4.9	428	35	AA549948	AA549948 1025m3 gm
C 29	50.2	4.9	578	35	AA550379	AA550379 1525m3 gm
C 30	50.2	4.9	644	40	AA942264	AA942264 LD26372.5
31	50	4.9	541	50	AU062000	AU062000 AU062000
32	49.8	4.9	314	20	T18187	T18187 0794c3 czap
33	49.8	4.9	536	38	AA440021	AA440021 LD15090.5
34	49.8	4.9	654	45	AU000948	AU000948 AU000948
C 35	49.8	4.9	551	45	AU000956	AU000956 AU000956
C 36	49.8	4.9	590	45	AU000995	AU000995 AU000995
C 37	49.8	4.9	573	45	AU001034	AU001034 AU001034
38	49.6	4.9	696	45	AU000518	AU000518 AU000518
39	49.4	4.8	529	26	W52352	W52352 zc48h03.s1
40	49.4	4.8	531	40	C90097	C90097 C90097 Dict
41	49.4	4.8	505	40	C90102	C90102 C90102 Dict
42	49.4	4.8	636	40	C92986	C92986 C92986 Dict
C 43	49.4	4.8	724	42	AI106800	AI106800 GH06024.5
44	49.4	4.8	681	44	AI260725	AI260725 LP04686.5
45	49.4	4.8	697	44	AI297514	AI297514 LP11826.5

ALIGNMENTS

RESULT 1
AU038896/c
LOCUS AU038896 610 bp mRNA
DEFINITION Dictyostelium discoideum SS (H.Urushihara) Dictyostelium
discoideum cDNA clone SSM114, mRNA sequence.
ACCESSION AU038896
NID 93985649
VERSION AU038896.1 GI:3985649

EST 29-MAR-1999


```

AI581441/c
LOCUS      AI581441      1145 bp      mRNA
DEFINITION SWOV3MCAM25C06SK Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmlL3) Onchocerca volvulus cDNA clone SWOV3MCAM25C06 5',
            mRNA sequence.
ACCESSION  AI581441
NID        G4567336
VERSION    AI581441.1 GI:4567336
KEYWORDS   EST.
SOURCE     Onchocerca volvulus.
ORGANISM   Onchocerca volvulus
            Eukaryota; Metazoa; Nematoda; Secernentea; Spiruria; Spirurida;
            Filarioidea; Onchocercidae; Onchocerca.
REFERENCE  1 (bases 1 to 1145)
AUTHORS    Williams,S.A., Lizotte-Waniewski,M., Laney,S. and Lustigman,S.
TITLE      Genes expressed in molting L3 larvae of Onchocerca volvulus
JOURNAL    Unpublished (1997)
COMMENT    On Mar 20, 1998 this sequence version replaced gi:2980257.

Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
            Location/Qualifiers
            1..1145
            /organism="Onchocerca volvulus"
            /strain="Kumba, Cameroons"
            /db_xref="taxon:6282"
            /map="16"
            /clone="SWOV3MCAM25C06"
            /clone_lib="Onchocerca volvulus molting L3 larva cDNA
            (SL96MLW-OvmlL3)"
            /dev_stage="molting L3"
            /lab_host="XL1-Blue MRF"
            /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
            Xho I; Filarial nematode parasite of humans. Third-stage
            larvae, L3, were isolated from infected black flies in
            Cameroons (forest strain). The L3 were cultured in 20% FCS
            in IMDM+ NCTC 135 and collected after day 1, 2, or 3 in
            culture. L3 of O. volvulus molt to fourth-stage larvae by
            day 5 in culture. mRNA was isolated from approximately
            6000 molting larvae (mL3), 2000 larvae from day 1, 2 or 3
            in culture, and converted to double-stranded cDNA using
            reverse transcriptase and oligo(dT) followed by RNase H
            and DNA pol I. The library was constructed in the lambda
            Uni-Zap XR vector and has 1 x 10E6 independent
            recombinants and the average insert size is ~1200 bp. The
            library was constructed by Sara Lustigman and Michelle
            Lizotte-Waniewski in the Laboratory of Dr. S. A. Williams.
            The library is available from Dr. Sara Lustigman (email:
            slustigm@nybc.org)."
```

```

BASE COUNT      320 a      106 c      77 g      468 t      174 others
ORIGIN
Query Match      5.2%; Score 52.6; DB 48; Length 1145;
Best Local Similarity 39.0%; Pred. No. 4.6;
Matches 270; Conservative 0; Mismatches 412; Indels 11; Gaps 3;

QY 324 TATTAAATGGTAATCTTGTAGTAATCAATATACACTTGCTAATGAATGAAGAAAAAGA 383
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1138 TATTAAAGGNTNATTTNGTANATANCATTAATTAANGTNGTNTCANANNNGNANAANTTAAAGA 1079
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 384 TATCAAAATTATGGTTAAATCAATGAGTTCAGGACTTGAATTTTATTCATTCACAAGGGAT 443
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1078 NATAGATAAAAATTAATTATNGATNANTAAATCTAAATACATAAAGAATCGANANAGNAA 1019
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 444 AATTCAATCGTGATATAAAACCCAGTAAATATTTTCTTCCCGGGATGATATAACACAAACC 503
```

```

Db 1018 ANTNTATNATNTAAGTAANTATATATANCAATTCATNGACAANNNAATAGTACGT 959
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 504 GATTATTGGAGATTTTGATATTTTGTATGATTTAA-AACTGCCACCTAAAGATGAACCCC 562
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 958 TAANCNGTTATATATAAAATATAATATATGATATAATTTANTGTCAATATANGTGTAATN 899
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 563 CTATGGCGAAATATATTGATGTATCTACAGGTATTTATAAAGCACCAGAAATGATCTTG 622
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 898 ATANTATGAANTATATANTTTATGTNATGANACAAANNANTNACATGATATANATAT-- 840
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 623 GTATACTAATTTATGAATATGAAATTTGATATTTTGGTCATTTGGGTATATATTTTACTGGTT 682
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 840 -NATCTGATNNTCANATANGATAANAGATNGTNGAGAGATANACATAAINTATNTAAGAA 782
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 683 TATATTCAGAAAAATTTTCAAAGTGTTTTGTAGTCAAAGATGATAAAGAAATGACTAATGATT 742
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 781 TATNNAATNTATNNAAAATNATNTANAANTTATATATATATNAATATTTATATGNATAANTTAN 722
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 743 CTCATGTTAGTGTATTTATATTTTAAATCAAATATTTGAAAAATTTCCGTACACCCAATT 802
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 721 ATATTATGTAANANTATAATAATTTATTAANNATTANTATATAATNTTAAANANAAN 662
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 803 TAACGTGATTTTGAAGATGAATTTTGTGTGATGAATATAA-----TAATGAAAACTTG 855
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 661 TNATTANNNTANNATTNAAANATTATNNGTANANATACAAGTAAAGATANTAAAAGATT 602
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 856 CATTTTAAAAAATTTCAATTTACAAAAATATCCTAGAAAAAGATTGGGATATATTTTACCT 915
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 601 ANTTTANATAAAATATAATTNCAAAAANTNCATANAATAATTTTANTATTANANTAACAANTAA 542
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 916 CGATGCAATGATGATTTAATGAAAGAAATTTTACCAAGATGATTAGATATGATCGAAGT 975
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 541 TNNAAAAATATTTAATATTTNNAAAAAAATCATATATAAAGTTTANTTAGAATNTTATNATG 482
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 976 AAAAGAATAACTTCTAAAGAAATCTTACAATTA 1008
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 481 NAATNTAATTTCAATATATANAATAANANANTTA 449
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
AI294319
LOCUS      AI294319      684 bp      mRNA
DEFINITION LP07659.5prime LP Drosophila melanogaster larval-early pupal POT2
            Drosophila melanogaster cDNA clone LP07659 5prime similar to
            U37541: Drosophila melanogaster complete mitochondrial genome, mRNA
            sequence.
ACCESSION  AI294319
NID        G3943726
VERSION    AI294319.1 GI:3943726
KEYWORDS   EST.
SOURCE     fruit fly.
ORGANISM   Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE  1 (bases 1 to 684)
AUTHORS    Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
TITLE      BDGP/HMI Drosophila EST Project
JOURNAL    Unpublished (1997)
COMMENT    On Jan 14, 1998 this sequence version replaced gi:1797824.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 76 row: E column: 11
High quality sequence stop: 623.
Location/Qualifiers
            1..684

FEATURES
source
```


Sat Nov 6 11:07:19 1999

us-09-072-994-13.rst

Page 10

Job time: 19381 sec

SEQUENCE CHARACTERISTICS:
LENGTH: 2483 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
US-08-946-617-1

Query Match 5.0%; Score 51.4; DB 3; Length 2483;
Best Local Similarity 45.5%; Pred. No. 0.013;
Matches 338; Conservative 0; Mismatches 396; Indels 9; Gaps 4;
Qy 281 AAATTACAAATATTGTAACGAACAACACGAGATTATTTATGGTATTAATGGTAACTTTG 340
Db 892 AAATTGTACAAGAAAGTATGAGAATAATATTAGAAATTATCTATAATAATAGTTTCTCTT 951
Qy 341 TTAGTAATCAATATACACTTGTCTAATGAATGAAGAAAAAGATATCAAAATTATGTTAA 400
Db 952 ATTATCTCATGGATTAGACCTAAGTTTCTTTTAAACAGCTATTATTCAATGTAATAA 1011
Qy 401 AATCAATGAGTTCAGGACTTGAATTTATTCATTCACAGGGGATAATTCATCGTGATATAA 460
Db 1012 ATTATATGCAACTGTAAATGATTTTATTAAGTAGATTAAATAAATGCTTTTGATACAA 1071
Qy 461 AACCCAGTAATATTTCTTTGCCCCGGGATGATATAACACACCGGATTATTTGGAGATTTG 520
Db 1072 TTCCACATAATATGTTAAT--AATGTAATAATGAGAGAATCAAGATAAAGGTTTCAT 1129
Qy 521 ATATTTGTTATGATTTTAAACTGCCACCTAAAGATGAACCCCTATGGCGAAATATATTG 580
Db 1130 AGACTTATTAT-ATAAATTATTAAGAGCTGGATATGTTGATAAATAATAATTATACATA 1188
Qy 581 ATGTATCTACAGGTATTTTATAAGCACCACCAATTTGATTTTGGGTATA--ACTAATATG 637
Db 1189 ATACAACCTTAGGAATTCCTCAAGGTAGTGTGTCAGTCTTATTTTATGTAATATTTT 1248
Qy 638 AATATGAAATGATATTTGGTCATTTGGGTATAATTTTGGTACTGGTTTATATTCAGAAAT- 697
Db 1249 TAGATAAATTAGATAAATAATTTAGAAAATAAATTTGAGAAATGAATTCATACTGGAATA 1308
Qy 697 -TTTCAAAGTGTTTTAGTCAAAGATGATAAAGAAATTTGACTAATGATTTCTCATGTTAGTG 754
Db 1309 TGTCTAATAGAGGTAGAAATCCCAATTTATAATAGTTTATCATCTAAATTTATAGATGA 1368
Qy 755 ATTTATATTATTAATCAATATTTTGAAATTTTGGGTACACCCAATTTAACTGATTTTG 814
Db 1369 AATTATTTATCTGAAAAATTTAAAAATTTGATTAGATTAAAGAGACCATTTACCAAGAAATATGG 1428
Qy 815 AAGATGAATATTTTGTGATGAATATAATGAAGAACTTGCAATTTTAAAAAATTCATTT 874
Db 1429 GATCTGATAAAGCTTTTAAAGAGCTTATTTTGTAGATATGCTGATGATATTTATCATTTG 1488
Qy 875 TACAAAAATATCCTAGAAAAGATTGGGATATTTATTTTACCTCGATGCAATGATGTTAA 934
Db 1489 GTGTAATGGGTTCTCATATATGATTTGTAATAATATTTTAAACGATATTTAATAACTTCTTAA 1548
Qy 935 TGAAGAAGAAATTTTACCAGATGATTAGATATGATCGAAGTAAAGATAAAGTCTTCTAAAG 994
Db 1549 AAGAAAATTTAGGTATGTCAATTAATATAGATAAATCCGTTATTAACATTTCTAAAGAG 1608
Qy 995 AAATCTTACAATTAATGTTGGAT 1017
Db 1609 GAGTTAGTTTTTTAGGGTATGAT 1631

RESULT 5
US-08-446-855A-1/c
; Sequence 1, Application US/08446855A
; Patent No. 5849573
; GENERAL INFORMATION:
; APPLICANT: Stewart, Thomas S

APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding carbamoyl
TITLE OF INVENTION: phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29.009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 4.9%; Score 50.4; DB 3; Length 8920;
Best Local Similarity 48.0%; Pred. No. 0.024;
Matches 144; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
Qy 102 TCTTAAATAGTTGATGAAGATTTTCAGTCTTCCACCACATTCATCCATCGAGAAATTT 161
Db 8824 TTTTAAAAAATCATACATTTTATAAATTATTAATAATTAAAGAAATTTATACATTT 8765
Qy 162 TATACTTAAACTTTGAAACACATCCAAACATAATTTGAATATTTTAAATGATCTTAAAT 221
Db 8764 AATGTTTATATATTTTATTTTATTAATAATAATAATTTATTAATAATATAAATAT 8705
Qy 222 TTATGATGATGTTATATTTAGTCACCAAAATTTGATCGTTATGATTTTGAGTCAATTTGA 281
Db 8704 TTGTAATTTGAAAAAACCAATTTTGGTTTATACATATGAGTGAATAAATAAATTTGTATA 8645
Qy 282 AATTACAAAAATTTGTAACGAACAACACGATTTTATTTATGTTATTTATGTTAATCTTGT 341
Db 8644 TATAACAAAAATTTTAAAAAATAAATAAATCAATATAAATAATGATCATATTTAATAAAA 8585
Qy 342 TAGTAATCAATATACACTTGTCTAATGAATTTGAAGAAAAAGATATCAATTTATGTTAAA 401
Db 8584 CATTTATTAATTTTAAACGTAATAATAATAATAATAATAATAATAATAATAATAATA 8525
RESULT 6
US-07-991-867B-8
; Sequence 8, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: NO. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66

TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1511 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entemopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (18..218)
FEATURE:
NAME/KEY: CDS
LOCATION: complement (234..782)
FEATURE:
NAME/KEY: CDS
LOCATION: 852..1511
US-08-107-755A-8

Query Match 4.8%; Score 49; DB 2; Length 1511;
Best Local Similarity 43.4%; Pred. No. 0.036;
Matches 325; Conservative 0; Mismatches 420; Indels 4; Gaps 2;
Qy 139 CATTCAATCCATCGAGAAATTTTATACCTTAAACTTTGAACACCATCCAAACATAATT 198
Db 761 CAATTATTAAATCTATCGACATTTTATTTATTTATTTGATATATTTTTCRAAAAAAAT 820
Qy 199 GAATATTTTAATGATCTTAAATTTTATGATGATGATGATATATAGTCACCAAAATGTATCGT 258
Db 821 TAATCAATGAAGAAAAAATAAATAATCAAAATGATTTACTAAATTTCTGATATAATTTT 880
Qy 259 TATGATTTGAGTCAATGATGAAATACAAATGAAATGTTAAACGAACAACGATTTATT 318
Db 881 AATAAATATTTTAAATATTTATTAATTTAAATAAATAAATAAATAAACAAGAGATAATGTTAT 940
Qy 319 TATGGTATTAATGGTAACTTGTAGTAATCAATATACACTTGC--TAATGAAATTTGAA 375
Db 941 TAATATTAATATATTTAAATAATTTAGTTAATTTAGAAAGAAATTCATATATATATATGA 1000
Qy 376 GAAAAGATATCAAAATTTATGTTAAATCAATGACTTCAGGACTTGAATTTATTCATTTCA 435
Db 1001 TAATAATATTTTAAATAATATTTCCAGAAATAATTTAAAGTTTATATATTTTCAAAATTTAA 1060
Qy 436 CAAGGGATTAATTCATCGTATATAAACCAGTAAATATTTTCTTTGCCCGGATGATATA 495
Db 1061 TATTATTATTTAAATTTTATACAAATTTAAATAATTAACATATTTAGATATATCTTA 1120
Qy 496 ACAC-AACCGATTATTCGAGATTTTGTATATTTTGTATGATTTTAAACTGCCACCTAAAGA 554
Db 1121 TAACAAAAATAGCAATATAAGTAATATATATATCTACACATTTCTATAGAAATTTTAAATTG 1180
Qy 555 TGAACCCCTATGGCGAAATATATTGATGATCTACAGGTATTTATAAAGCACCAGAATT 614
Db 1181 TGAATCATGTAATATAAATGACTATATTTTATTAATAATTTAGTAATTTTAAATAAATTT 1240
Qy 615 GATTCCTTGGTATAACTAATTTATGAATGAAATGATATTTTGGTCATTTGGGTATATTTT 674
Db 1241 AATAATATCTAAAAATAAATTTTGGTAACCTTTAATAATGTTTTTCTATTAGTATAGTTGA 1300
Qy 675 GACTGGTTTATATTCAGAAAAATTTTCAAAGTGTTTTAGTCAAGATGATATAAACAATTGAC 734
Db 1301 GTTAAATATGGAATCAATACAAATAAAGATTTATAAATTTATAGAAAAATTTAATTT 1360
Qy 735 TAATGATTTCTCATGTTAGTATTTATATTTTATTAATCAATATTTTGAAAAATTTCCGGTAC 794
Db 1361 AAAAAAATTAGATATATCTTTCAATGTTAAAAAATAAATAATATATACATTTTGATAAAATTTCC 1420
Qy 795 ACCCAATTTAACTGATTTTGAAGATGAATTTTGTGTGATGAATATATATATGAATACTT 854

Db 1421 AAAAAAGTATAAATCAATTTATGTATATCAATCATATAAAGAAATTTATAATTTATAA 1480
Qy 855 GCATTTTAAAAAATTCATTTACAAAAT 883
Db 1481 AAATTTATCAAAATATAATTTGAATATGAAT 1509
RESULT 8
US-08-544-332-8
; Sequence 8, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: NO. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UFI14.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1511 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Amsacta moorei entemopoxvirus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (18..218)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (234..782)
; FEATURE:
; NAME/KEY: CDS

LOCATION: 852...1511
US-08-544-332-8

Query Match 4.8%; Score 49; DB 4; Length 1511;
Best Local Similarity 43.4%; Pred. No. 0.036;
Matches 325; Conservative 0; Mismatches 420; Indels 4; Gaps 2;

QY 139 CATCAATCCATCCAGAAATTTTATATCTTAAACCTTGAAACCATCCCAACATAATT 198
DB 761 CAATATTAATCTATCGACATTTTATATCTTAAACCTTGAAACCATCCCAACATAATT 820
QY 199 GAATATTTTAAATGATCTTAAATTTTATGATGATGTTATATTAGTCACCAAAATGTATCGT 258
DB 821 TAATCAATGAAAAAATAAATAAATTTATCAAAATGGATTTACTAAATCTGTGATATAATTTT 880
QY 259 TATGATTTGAGTCAATTTGATGAAATTTACAAAAATATTGTAAACGCAACACACGATTTATT 318
DB 881 AATAAATATTTTAAATATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 940
QY 319 TATGGTATTAATGGTAATCTTTAGTAAATCAATATACACTTGC---TAATGAAATTTGAA 375
DB 941 TAATATTAATATTAATAAATAAATTTAGTTAAATTTAGAGAATTTGCATATATATATATGA 1000
QY 376 GAAAAAGATATCAAAATTTATGGTTAAATTTCAATGAGTTTCAAGGACCTGAATTTTATCATTTCA 435
DB 1001 TAATAATATTTTAAATAATATTTCCAGAAAAATATTAAAGTTTATATATTTTCAAAATTTAA 1060
QY 436 CAAGGGATAATTCATCGTGATATATAAACCAGTAATTTTCTTGGCCCGGATGATATA 495
DB 1061 TATATTAATTTAAATTTTATACAAAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1120
QY 496 ACAC-AACCGATTATGGAGATTTTGGATTTTGTATGATTTTAAACTGCCACCTAAAGA 554
DB 1121 TACAAAAAATAGCAATATAGTAATATTTATACCTACCACTCTATAGAAATTTTAAATTTG 1180
QY 555 TGAACCCCTATGGCGAAATATATTGATGATCTACAGGTATTTATATAAGCACCAGAAAT 614
DB 1181 TGAATCATGTAATATAATGACTATATATTTTATTAATTAATTTAGTAAATTTTAAATTTTAA 1240
QY 615 GATTCCTGGTAACTAATTAATGAATTAATGAATTTGATTTTGGCATTTGGGTATATAATTTT 674
DB 1241 AATATATCTAAAAATAAATTTGGTAACTTAATATGTTTTCCTATTAGTATAGTTGA 1300
QY 675 GACTGGTTTATATTCAGAAAAATTTTCAAAAGTGTTTTACTCAAGATGATAAAGAAATTTGAC 734
DB 1301 GTAAATATGGAATCAATACAAATAAAGATATATAATTTATAGAAAAATTTAATTAATTT 1360
QY 735 TAATGATTTCTGTTAGTATTTATATTTATTAATCAATATTTGAAAAATTTCCGGTAC 794
DB 1361 AAAAAAATTAGATATATCTTTCAATGTTTAAAAAATAATATACATTTGTATATAATTTCC 1420
QY 795 ACCCAATTTAACTGATTTTGAAGATGAATTTTGTGATGAATATAATTAATGAACATTT 854
DB 1421 AAAAGTATAACTCATTTTATGTTATATCAATCATATAAAGAAATTTAATTAATTTTAA 1480
QY 855 GCATTTTAAAAAATTTCAATTTACAAAAAT 883
DB 1481 AAATTTATCAATATAATTTGAATATGAAT 1509

RESULT 9
US-07-991-867B-1/c
; Sequence 1, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: NO. 5476781el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: FL
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B
FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8457 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (65..1459)
FEATURE:
NAME/KEY: CDS
LOCATION: 1474..2151
FEATURE:
NAME/KEY: CDS
LOCATION: complement (2239..2475)
FEATURE:
NAME/KEY: CDS
LOCATION: 2502..2987
FEATURE:
NAME/KEY: CDS
LOCATION: 3080..6091
FEATURE:
NAME/KEY: CDS
LOCATION: complement (6277..6768)
US-07-991-867B-1

Query Match 4.8%; Score 48.8; DB 1; Length 8457;
Best Local Similarity 47.7%; Pred. No. 0.05;
Matches 207; Conservative 0; Mismatches 222; Indels 5; Gaps 2;
QY 475 TTCTTTGCCCGGATGATATACACACCGATTTATTTGGAGATTTTGGATATTTTGTATGAT 534
DB 614 TTATATACAGAACTTATAAAACAAAAATAAATAATATATGTTTATTGAATCTGATATGAC 555
QY 535 TAAAACTGCCACCTAAAGATGAACCCCTATGCGGAAATATATTGATGATCTACAGGT 594
DB 554 TATAAAAGCCA--TAAACACTACATAGTACTCTTAATACTAATATTGCTTTATTATAC 497
QY 595 ATTATAAAGCACCAGAAATTTGTTCTTGGTATACTAATTAATTAATGAATGAATTTGATATT 654

Db 496 ATTGATGACTATCAGAAATATTAAATGATATATTAAATGTAATATTACTTTTT 437
QY 655 TGGTCATTGGGTAAATTTTGGCTGGTTTATATTCAGAAAATTTTCAAGTGTGTAGTC 714
Db 436 AATACATTTGAATTACAATATGATAGTCCCGATTGTTGATGTTTAAATATAATTTTA 377
QY 715 AAAGATGATAAAGAAATGACTAATGATTCATGTTAGTGCATTTATATTTAAATCAA 774
Db 376 TATTATATAGTTTATTTCACATAAAATCTAAATTTGAATTTAAATAATATTATTCT 317
QY 775 ATATTGAAAATTTCCGGTACACCCCAATTTAACTGATTTTGAAGATGAATTTTGTGAT 834
Db 316 ATGAGTTTATTGGTGATTTATTGGCT---AGTAGTTATAGAGGTCGATTATTATTCT 260
QY 835 GAATATAAATGAAGAACTTGCATTTTAAATAATTCATTTACAAAATATCCTAGAAA 894
Db 259 AGATATGATATTATAGTATTGACGAATTTTAAATAACATACATTAGAATATTAAATATA 200
QY 895 GATTGGGATATTAT 908
Db 199 AATAAAAGTTTAT 186

RESULT 10
US-07-991-867B-25/C
; Sequence 25, Application US/07991867B
; Patent No. 5476781
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5476781e1 Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/991,867B
FILING DATE: 12-DEC-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 92/14818
FILING DATE: 12-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,685
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1395 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)

US-07-991-867B-25
Query Match 4.8%; Score 48.8; DB 1; Length 1395;
Best Local Similarity 47.7%; Pred. No. 0.039;
Matches 207; Conservative 0; Mismatches 222; Indels 5; Gaps 2;
QY 475 TTCTTTGCCGGGATGATATAACACACCGGATVATTTGGAGATTTTGATATTGTTATGAT 534
Db 550 TTATATACCAGAACTTATAAAACAAATAAATAATATATGTTTATTGAATCTGATATGAC 491
QY 535 TTAAGACTGCCACCTAAAGATGAACCCCTATGGCGAAATATATTGATGATCTACAGGT 594
Db 490 TATAAAAGCCA--TAAACACTACAATAGTACTCTTAATACTAATTAATGCTTATTATAC 433
QY 595 ATTTATAAGCACCAAGATGATTTGTTGGTATTAATACTAATTAATGAATATGAATTTGATATT 654
Db 432 ATTGATGTACTATCAGATATTTTAAATGATATATTTAAATGTAATTTACTTTTTT 373
QY 655 TGGTCATTTGGGTATATTTTGGCTGTTTATATTCAGAAAATTTTCAAGTGTTTTAGTC 714
Db 372 AATACTTTTGAATTTACAATATGATAGTCCCGATTGGTATGTTTAAATAATATTTTTA 313
QY 715 AAAGATGATAAAGAAATGACTAATGATTTCTCACTGTTAGTGTATTTATTTTAAATCAA 774
Db 312 TATTATATAGTTTATTTCACATATAAATCTAAATTTGAATTTAAATAATATTATTCT 253
QY 775 ATATTGAAAATTTCCGGTACACCCCAATTTAACTGATTTTGAAGATGAATTTTGTGAT 834
Db 252 ATGAGTTTATTGGTGATTTATTGGCT--AGTAGTTATAGAGGTCGATTATTATTCT 196
QY 835 GAATATAAATGAAGAACTTGCATTTTAAATAATTTCAATTTACAAAATATCCTAGAAA 894
Db 195 AGATATGATATTATAGTATTGACGAATTTTAAATAACATACATTAGAATATTAAATATA 136
QY 895 GATTGGGATATTAT 908
Db 135 AATAAAAGTTTAT 122

RESULT 11
US-08-107-755A-1/C
; Sequence 1, Application US/08107755A
; Patent No. 5721352
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5721352e1 Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: U.S.A.
; ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,755A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,658
FILING DATE: 30-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C2
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6768 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Amsacta moorei entomopoxvirus
FEATURE:
NAME/KEY: CDS
LOCATION: complement (65..1459)
FEATURE:
NAME/KEY: CDS
LOCATION: 1474..2151
FEATURE:
NAME/KEY: CDS
LOCATION: complement (2239..2475)
FEATURE:
NAME/KEY: CDS
LOCATION: 2502..2987
FEATURE:
NAME/KEY: CDS
LOCATION: 3080..6091
FEATURE:
NAME/KEY: CDS
LOCATION: complement (6277..6768)
US-08-107-755A-1

Query Match 4.8%; Score 48.8; DB 2; Length 6768;
Best Local Similarity 47.7%; Pred. No. 0.048;
Matches 207; Conservative 0; Mismatches 222; Indels 5; Gaps 2;
QY 475 TTCTTTGCCCGGATGATATACACACACCGATTATTGGAGATTTTGATATTTGTTATGAT 534
Db 614 TTATATACAGAACTATATAACACAAATAAAATATATGTTTATTGAATCTGATATGAC 555
QY 535 TTAAACATGCCACCTAAAGATGAACCCCTATGGCGAAATATATTGATGATCTACAGGT 594
Db 554 TATAAAGGCCA--TAAACACTACAATAGTACTCTTAATCTAATTAATGCTTATTATAC 497
QY 595 ATTTATAAGCACCAGAAATTGATCTTGGTATACTAATTAATTAATTAATTAATTAATTA 654
Db 496 ATTGATGTACTATCAGAAATTTAAATGATATATTTAAATGTAATTAATTAATTAATTTT 437
QY 655 TGGTCATTGGGTATAAATTTTGACTGGTTTATATTCAGAAAATTTTCAAAGTGTTTATC 714
Db 436 AATACCTTTGAATTACAATATGATAGTCCCGATTGGTATGTTTAAATTAATTAATTTTA 377
QY 715 AAAGATGATAAAGAAATTGACTAATGATCTCATGTTAGTGATTTATATTATTAATTA 774
Db 376 TATTATATAGTTTATTTCACATAAAATCTAAATTTGAATTTAAATAATTAATTAATCT 317
QY 775 ATATTTGAAAATTCGGTACACCCAAATTTAACTGATTTTGAAGATGAATTTTGTGAT 834
Db 316 ATGAGTTTATTGGTGATTTTATGGCT--AGTAGTATAGAGTGCGATTTATTCT 260
QY 835 GAATATAAATGAACCTTGCAATTTTAAATAATTTCAATTTTACAAAAATTCCTAGAAA 894
Db 259 AGATGATATATTATGATTGACGAAATTAATAATACATTAGAAATTTAATATAAAA 200
QY 895 GATTGGGATATTAT 908
Db 199 AATAAAGTTTAT 186

RESULT 12
US-08-107-755A-25/c
Sequence 25, Application US/08107755A
Patent No. 5721352
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Hall, Richard L.
APPLICANT: Guidl, Michael E.
TITLE OF INVENTION: No. 5721352el Entomopoxvirus Expression System
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/107,755A
FILING DATE: 19-AUG-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/827,658
FILING DATE: 30-JAN-1992
APPLICATION NUMBER: US 07/657,584
FILING DATE: 19-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF114.C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (904) 375-8100
TELEFAX: (904) 372-5800
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1395 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-107-755A-25

Query Match 4.8%; Score 48.8; DB 2; Length 1395;
Best Local Similarity 47.7%; Pred. No. 0.039;
Matches 207; Conservative 0; Mismatches 222; Indels 5; Gaps 2;
QY 475 TTCTTTGCCCGGATGATATACACACACCGATTATTGGAGATTTTGATATTTGTTATGAT 534
Db 550 TTATATACAGAACTATATAACACAAATAAAATATATGTTTATTGAATCTGATATGAC 491
QY 535 TTAAACATGCCACCTAAAGATGAACCCCTATGGCGAAATATATTGATGATCTACAGGT 594
Db 490 TATAAAGGCCA--TAAACACTACAATAGTACTCTTAATTAATTAATTAATTAATCT 433
QY 595 ATTTATAAGCACCAGAAATTGATCTTGGTATACTAATTAATTAATTAATTAATTAATTA 654
Db 432 ATTGATGTACTATCAGAAATTTAAATGATATATTTAAATAATGTAATTAATTAATTTT 373
QY 655 TGGTCATTGGGTATATTTTGACTGGTTTATATTCAGAAAATTTTCAAAGTGTTTATC 714
Db 372 AATACCTTTGAATTACAATATGATAGTCCCGATTGGTATGTTTAAATTAATTAATTTTA 313
QY 715 AAAGATGATAAAGAAATTGACTAATGATCTCATGTTAGTGATTTATTAATTAATTA 774
Db 312 TATTATATAGTTTATTTCACATAAAATCTAAATTTGAATTTAATAATTAATTAATTTCT 253

[illegible]

RESULT 13
US-08-544-332-1/c
; Sequence 1, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777e1 Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77

```

;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1474..2151
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (2239..2475)
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2502..2987
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3080..6091
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: complement (6277..6768)
;
; US-08-544-332-1

```

	Query Match	4.8%;	Score 48.8;	DB 4;	Length 8457;
	Best Local Similarity	47.7%;	Pred. No. 0.05;		
	Matches 207;	Conservative	0; Mismatches 222;	Indels 5;	Gaps
QY	475	TTCTTTGCCCGGATGATATAACACAACCGATTATTGGAGATTTTGATATTTGTATTATGAT	534		
Db	614	TTATATACCAGAACTTATAAACACAAATAAAAAATATATGTTTAITGAATCIGATATGAC	555		
QY	535	TTAAAAC TGCCACCTAAGAGTGAACCCCCTATGGCGAATATATTGATGATCTACAGGT	594		
Db	554	TATAAAAGCCA--TAAACACTACAATAGTACTCCTATACTAAATATGCTATTATTATAC	497		
QY	595	ATTTATAAGGCAC CAGAATTGATTC TTGGTATAACTAAATATGATGAATTAATGAAATGATAIT	654		
Db	496	ATTGATGTACTATCAGAAATTTAAATGATATATTAAAAGTGAATAATATTACTTTTTTT	437		
QY	655	TGGCATTTGGGTATAATTTTGACTGGTTTATATTCAGAAAATTTTCAAAGTGTTTTAGTC	714		
Db	436	AATACTTTTGAATTACAATATGATAGTCCCCGATGGGTATGGTATGTTTAATATATATTTTA	377		
QY	715	AAAGATGATAAGAATTGACTAATGATTC TCA GTTAGT GATTTATATTTAATAATCAA	774		
Db	376	TATTATATAGTTTATTTCACATAAAATCTRAATTTGAATTTAAAAAATATATATTCT	317		
QY	775	ATATTGAAAATTCGGFRACACCCAAATTTA ACTGATTTTGAAGATGAATATTTTGTGAT	834		
Db	316	ATGAGTTTTATTGGTGATTATTGGCT--AGTAGTTATAGAGGTGCATTATTATTCT	260		
QY	835	GAATATAATATGAAAACPTGCATTTTAAAAAATTC AATTTACAAAATATCCTAGAAAA	894		
Db	259	AGATATGATATTAAATAGTATTGACGAATTTAAAAAATACATTAGAAATATTTTAATATAAAA	200		
QY	895	GATTGGGATATTAT	908		
Db	199	AATAAAAAAGTTTAT	186		

```

RESULT 14
US-08-544-332-25/c
; Sequence 25, Application US/08544332
; Patent No. 5935777
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Hall, Richard L.
; APPLICANT: Gruidl, Michael E.
; TITLE OF INVENTION: No. 5935777el Entomopoxvirus Expression System
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Gerard H. Bencen
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:

```


; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/544,332
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/991,867
; FILING DATE: 07-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,755
; FILING DATE: 19-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 92/14818
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,685
; FILING DATE: 30-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/657,584
; FILING DATE: 19-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Bencen, Gerard H.
; REGISTRATION NUMBER: 35,746
; REFERENCE/DOCKET NUMBER: UFI14.C4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 904-375-8100
; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-544-332-25

Query Match 4.8%; Score 48.8; DB 4; Length 1395;
Best Local Similarity 47.7%; Pred. No. 0.039;
Matches 207; Conservative 0; Mismatches 222; Indels 5; Gaps 2;

QY 475 TTCTTTCCCGGATGATATAACACACCGATTATTGGAGATTGTGATATTTGTTATGAT 534
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 TTATATACCAGAACCTATAAAACAAATAAATAATATATGTTTATGAATCTGATATGAC 491
QY 535 TTAACACCTGACCACTAAAGATGAACCCCTATGCGGAATATATGATGATCTACAGGT 594
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
490 TATAAAGCCCA--TAAACACTACATAGTACTCTCTAATACATAATTATGCTATTATAC 433
QY 595 ATTTATAAGCACCAGAAATTGATTTGGTATACATAATTAATGATGATGAATGATATT 654
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
432 ATGATGCTACTATCAGAAATATTTAAATGATATATTTAAATAATGATAATTTACTTTTT 373
QY 655 TGGTCATTGGGTATAATTTTGACTGGTTTATATTCAGAAAAATTTCAAAGTGTGTTAGTC 714
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
372 AATACITTTGAATTACAATATGATATGATAGTCCGATTGTTGTTAATATATATTTTAA 313
QY 715 AAAGATGAIAAAGAAATGACTAATGTTCTCATGTTAGTAGTATATATTTTAAATCAA 774
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
312 TATTATATAGTTTATTCAACATAAATCTAAATTTGAATTTAAAAAATTATATTCTTCT 253
QY 775 ATATTTGAAAATTTTCGGTACACCCCAATTTAACTGATTTTGAAGATGAATTTTGTGAT 834
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
252 ATGAGTTTTATTGGTGATTTATTGGCT--AGTAGTATAGAGGTGATTTATTCTTCT 196
QY 835 GAATATAATAAATGAAAACCTGCAATTTTAAAAAATTCAAATTTACAAAAATATCTAGAAA 894
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 AGATATGATATTAATAGTATTTGACGAATTTAAAAATACATTAGAAATATTTAATATAAAA 136
QY 895 GATTGGGATATTAT 908

Db 135 AATAAAAGTTTAT 122
|| | ||||
RESULT 15
US-08-480-604A-9
; Sequence 9, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7101 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..7098
; US-08-480-604A-9

Query Match 4.7%; Score 48; DB 2; Length 7101;
Best Local Similarity 43.8%; Pred. No. 0.07;

Matches	310;	Conservative	0;	Mismatches	390;	Indels	8;	Gaps	2;	
QY	195	AATTGAATATTTTAAATGATCTTAAATTTTATGATGATGTTATATTTAGTCACCAAAATTTGTA								254
Db	6105	ATTTAAATATTTTGTCTCATCATATAATGAAGATTTAGGAAATGAAGAAAGGTGAAGAAATCTC								6164
QY	255	TCGTTATGATTGAGTCAATGATTGAAATTAACAAAATATTTGTAACGAACAACACGATT								314
Db	6165	ATATTCGGTATATTAATTTCAATAATAAAAATTTACTATTTTGTGATTCATTACAGC								6224
QY	315	TATTTATGGTATTAAATGGTAATCTTGTGTAGTAATCAATATACACTTGCTAAATGAATGA								374
Db	6225	TGTAGTTGGATGGAAGATTTAGAGGATGGTTCAAAGTATTATTTTGTGATGAAGATACAGC								6284
QY	375	AGAAAAAGATATCAAAATTTATGGTTAAAAATCAATGAGTTCAGGACTTGAATTTTATTCATTC								434
Db	6285	AGAAGCATATATAGGTTTGTTCATTAATAAATGATGGTCAATATTAATTTTAAATGATGATGG								6344
QY	435	ACAAGGGATAATTCATCGTGATATAAAACCCAGTAATATTTTCTTTGCCCGGGATGATAT								494
Db	6345	AATTATGCAAGTTGGATTTGTCACTATAAATGATAAAGTCTTCTACTTCTCTGACTCTGG								6404
QY	495	AACACAAACCGAATATTGGAGATTTTGATATTTTGTATGATTTTAAACTGCCACCCTAAAGA								554
Db	6405	AATTATAGAATCTGGAGTACAAAACATAGATGACAATTTATTTCTATATAGATGATAATGG								6464
QY	555	TGAACCCCTATGGCGAAATATATTGATGTATCTACAGGTATTTATATAAGCACACAGAATT								614
Db	6465	TATAGTTCAAATTTGGTGTATTTGATACTTCAGATGGATATAAATATTTTGCACCTGCTAA								6524
QY	615	GATTCCTGGTATAACTAATTAATGAATATGAATTTGATATTTTGGTCATTTGGGTATAATTT								674
Db	6525	TACTGTAATGATAATATTTACGGACACAGTCAGTTGAATATAGTGGTTTAGTTAGAGTTGG								6584
QY	675	GACTGGTTTATATTCAGAAAAATTTTCAAAGTGTTTTAGTCAAAGATGATAAAGAATTGAC								734
Db	6585	GGAAGATGTATATT-----ATTTTGGAGAAACATATACAAATTGAGACTGGATGGATATA								6638
QY	735	TAATGATCTCATGTTAGTGATTTTATATTTTATAAATCAAAATATTTGAAAAATTTCCGGTAC								794
Db	6639	TGATATGGAATAATGAAAGTGATAAATATTTATTCATCCAGAACTAAAAAAGCATGCAA								6698
QY	795	ACCCA--ATTTAACTGATTTTGAAGATGAATTTATTTTGTGATGAATATATAATGAAAAC								852
Db	6699	AGGTATTAATTTAATTGATGATATAAATAATTAATTTTGTGATGAGAGGGCATATAATGAGAAC								6758
QY	853	TTGCATTTTAAAAAATTCAAATTTACAAAAATATCCAGAAAAAGATTGG								900
Db	6759	GGGTCTTATATCATTTGAAATAATAAATTAATTAATTTACTTTAATGAGAAATGG								6806